R E P O R T R E S U M E S

ED 019 713 24 CG 002 091

A STUDY OF METHODS OF ORGANIZING LEARNING. FINAL REPORT.

BY- SEIBEL, ROBERT
PENNSYLVANIA STATE UNIV., UNIVERSITY PARK

REPORT NUMBER BR-5-8098 PUB DATE DEC 67
REPORT NUMBER CRP-S-297
CONTRACT OEC-5-10-431
EDRS PRICE MF-$0.50 HC-$2.68 65P.

DESCRIPTORS- *LEARNING PROCESSES, EXPERIMENTS, *COGNITIVE PROCESSES, TASK PERFORMANCE, *CLASSIFICATION, LOGICAL THINKING, TAXONOMY, *ASSOCIATIVE LEARNING,

GIVEN THE TASK OF LEARNING A SERIES OF RANDOMLY ORDERED MEANINGFUL ITEMS, MOST SUBJECTS (SS) IMPOSE SOME ORGANIZATION ON THE MATERIAL DURING THE PROCESS OF LEARNING. AN EXPERIMENTAL PARADIGM IS DESCRIBED WHICH PERMITS THE OBJECTIVE SCORING OF EACH S'S SUBJECTIVE ORGANIZATION OF THE MATERIAL ON EACH LEARNING TRIAL. THE CENTRAL FEATURE OF THE PARADIGM IS THE UTILIZATION OF AN ESSENTIALLY UNSTRUCTURED STUDY SHEET WHICH EACH S PREPARES ANEW, FOR HIS OWN USE, DURING EACH LEARNING TRIAL. NEW INDICES ARE DEVELOPED FOR MEASURING--(1) THE CONSISTENCY OF ORGANIZATION (INDEPENDENT OF SEQUENTIAL ORDER) FROM ONE TRIAL TO THE NEXT, (2) THE STEREOTYPY OF ORGANIZATION ACROSS SS ON ANY ONE TRIAL, AND (3) THE EXTENT TO WHICH SPECIFIC PAIRS OF WORDS ARE TO BE FOUND IN THE SAME SUBJECTIVE CATEGORIES FOR DIFFERENT SS. IN ADDITION TO THE STUDY SHEET PARADIGM, TWO EXPERIMENTS EMPLOYING THAT PARADIGM, AND SEVERAL NEWLY DEVELOPED MEASURES FOR PREVIOUSLY UNMEASURED ASPECTS OF THE SUBJECTIVE ORGANIZATION OF MATERIAL ARE DESCRIBED. A THIRD EXPERIMENT IS CONCERNED WITH A PARTICULAR IMPLICATION OF SUBJECTIVE ORGANIZATION BEHAVIOR. IT WAS FOUND THAT IN A MIXED LIST OF ITEMS, EXHAUSTIVE CATEGORIES ARE LEARNED FASTER THAN ARE EXHAUSTIVE CATEGORIES WITH ONE ITEM MISSING. (AUTHOR/IM)

ED019713

BR-5-8098
P.A.24

FINAL REPORT 24

Project No. 5-8098-2-12-1(S-297)

Contract No. OE-5-10-431

A STUDY OF METHODS OF ORGANIZING LEARNING

December 1967

U. S. DEPARTMENT OF

HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

CG 002 091

A STUDY OF METHODS OF ORGANIZING LEARNING

Project No. 5-8098-2-12-1(S-297)
Contract No. OE-5-10-431

Robert Seibel

December 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

The Pennsylvania State University

University Park, Pennsylvania

ACKNOWLEDGEMENTS

The author wishes to acknowledge the initial support of Warren H. Teichner, and the assistance and contributions of Richard Olsen and William Amatneek.

CONTENTS

	Page
Acknowledgments	
Introduction	1
The Experimental Paradigm	1
The Three Experiments	2
Experiment One	2
Method	2
Results	3
Discussion	13
Experiment Two	14
Method	14
Results	17
New Measures of Subjective Clustering	27
Results of Experiment Two for New Measures	30
Discussion	34
Experiment Three	36
Method	38
Results	38
Discussion	40
Conclusions and Implications	41
Summary	43
References	45
Appendix A	A-1
Appendix B	B-1

Introduction

Attention to the behavior of individual subjects (Ss) certainly is not a new idea, but despite arguments for it, group experiments have predominated in the verbal learning area. Further, about the only experiments that have been concerned with structure and patterning in learning have been those concerned with clustering in free recall. But these experiments have also been primarily concerned with group performance. Typically, they have manipulated aspects of the stimuli in terms of prior norms generated by prior groups of Ss and then scored the behavior of the experimental Ss in terms of whether it did or did not conform to what the experimenter (E) had "built-in" via the norms. A concept attainment view as differentiated from a concept formation one (Bruner, et. al., 1957, pp. 21-22 and 44). Clustering on the part of Ss which did not conform to that which E had built in was usually ignored or at best mentioned in passing. Until very recently the only studies directly concerned with subjective organizations in learning have been those of Tulving (1962), and of Carterette and Coleman (1963). These studies, however, have been limited to examination of the degree to which Ss report words in the same order from trial to trial (by means of a measure called SO), and have been completely insensitive to clustering unless it was accompanied by relatively rigid sequencing.

Another approach to the analysis of subjective organization is exemplified in a study recently reported by Marshall (1967). This study classified clustering into two kinds: a) experimenter-defined, and b) idiosyncratic, but did so by means of a post-experiment recognition association test. Mandler (1967) reports a series of studies in which S subjectively organized the material prior to being tested for learning. The subjective clustering of individual Ss during learning has not been studied, except in terms of the SO (or a closely related) measure.

The present report describes an experimental paradigm designed to measure subjective clustering of individual Ss during learning, two experiments employing that paradigm, and a third experiment concerned with a particular implication of subjective organization behavior.

The Experimental Paradigm

The basic procedural elements of the paradigm are the following:
1. A presentation period during which items to be memorized are presented one at a time (five sec. per word in the present studies), in a random (or pseudo-random) order. During this period the Ss are required to write each item down as they see (or hear) it on a specially prepared

study sheet. This study sheet contains nothing but an array of blank cells (e.g., on an 8 1/2 " x 14" sheet of paper with a matrix of 12 columns and 28 rows). Each word may be written in any cell in the array but only one word per cell. 2. (After completion of the presentation.) A study period (one to one and a half minutes in the present experiments) during which the S is allowed to study his personally created study sheet. 3. A test period (a single block of time equivalent to four to four and a half sec. per word in the present experiments) during which the S is instructed to write the words in a list in the order in which they occur to him. 4. Repeats of the preceding three steps (in the present experiments either four or five repeats). Each repeat involves a new random order of presentation and new study and test sheets, old sheets having been removed at the end of the appropriate periods. Current experiments also incorporate a one and one half minute "pre-look" at a randomly arranged simultaneous presentation of all of the words to be memorized (words randomly arranged on a study sheet). This "pre-look" precedes the first presentation only. Instructions to the Ss suggest no particular mode of organization on their study sheets, but do say "arrange the words on the study sheet to best help you memorize."

The first experiment also included a post-experiment period during which each S indicated (by bracketing and labeling) how he attempted to organize the items and why. This last step was solely for checking on the "validity" of the objective scoring procedures applied to the subjective organizations on the study sheets. With rectangular arrays of cells on the study sheets, Ss employ horizontal or vertical (rarely mixed) lists on their study sheets. Within these overall orientations (which may be established objectively, cf. Experiment Two) simple adjacency on the study sheet is an adequate (and "valid") criterion for defining "belonging to the same organizational unit." In other words, if an S organizes vertically, e.g., then each column of adjacent words in the matrix contains a cluster so far as the S is concerned. Thus, subjective clusters may be objectively defined on the basis of performance during the learning experiment.

The Three Experiments

Experiment One

This was the first test of the paradigm. It was anticipated that the paradigm might provide for a "look inside the S" in what was otherwise a typical free-recall experiment. Hence, a control group (Control 2) was included that learned under "typical free-recall" conditions. Another control group (Control 1) was also included to check on the possible effects of the overt "organizing" activity expected of Ss using the study sheets.

Method for Experiment One. Three groups of Ss each learned forty words. Thirty-four of the words were taken from Underwood and Richardson's norms (1956) and consisted of four categories of high dominance

and four categories of low dominance words, and four words in a miscellaneous category such that there were minimal relations amongst those words and between those words and the eight other dominance categories. The remaining six words on the list were in two categories, namely, table utensils and class days. Table 1 presents the forty words, their dominance classification, and category names.

- - - - - - - - - - - - - - -
Insert Table 1 about here
- - - - - - - - - - - - - - -

The experimental (n = 24) group learned via the paradigm as described. Timing was such that there were five sec. per word during presentation (via slides), there were sixty sec. of study during which the S studied his personally created study sheet, and there were 180 sec. (4 1/2 sec. per word) for writing on the tests. There were four repetitions of the sequence, or four trials. The Ss of the first control group, Control 1 (n = 24), had exactly the same conditions except for their study sheets. Their study sheets were just a single column of 40 spaces and they were instructed to write the words on their study sheets in order as they saw them. Thus the only difference was that they did not have an opportunity to organize on their study sheets. The S of a second control group, Control 2 (n = 21), had essentially the same conditions except that they did not have a study sheet at all, nor did they have the sixty sec. of study time. They were given five trials, and for comparing this group with the others their performance (words correct and a clustering score) was linearly interpolated so that comparisons were made at four points at which the three groups had equal times in the learning situation.

Results for Experiment One The overall results in terms of number of words correct are depicted in Figure 1. An analysis of variance indicates

- - - - - - - - - - - - - - -
Insert Figure 1 about here
- - - - - - - - - - - - - - -

a significant difference amongst groups ($F = 8.171$, $df = 2,66$, $p < 0.001$), a significant trials effect ($F = 473.02$, $df = 3,198$, $p < 0.001$), and a non-significant trials-by-groups interaction ($F = 1.376$, $df = 6,198$, $0.20 < p < 0.25$). Collapsing across trials and applying the Newman-Keuls method of *a posteriori* comparisons (Winer, 1962) to the resultant analysis of variance indicates a significant difference ($p < 0.01$) between the experimental group and each of the two control groups, but no significant difference ($p > 0.05$) between the two control groups. The means for total number of words correct are: Exper. grp. = 131.04, Control 1 = 119.25, and Control 2 = 119.89.

The 24 Ss in the experimental group divided themselves as follows: six of them wrote the words on their study sheets in the order in which they were presented, i.e., they did not organize (sub-group NO) on their study sheets. Five of the Ss alphabetized (sub-group AO), and the

TABLE I. The forty words used in Experiment One, classified according to dominance level (Underwood and Richardson, 1956) and category.

High Dominance (> 75)		Low Dominance		Miscellaneous and "cue"	
atom, crumb, flea, germ	small	belly, derby, dime, platter	round	ski, lard, puddle, paste	not related
ammonia, garbage, manure, skunk	smelly	camel, chestnut, cinnamon, freckle	brown	knife, fork, spoon	table utensils
zoo		earthworm, jellyfish, lizard	slimy	Monday, Wednesday, Friday	class days
auditorium, mansion, stadium, whale	big	badge, bracelet, buckle	shiny		
canary, dandelion, mustard	yellow				

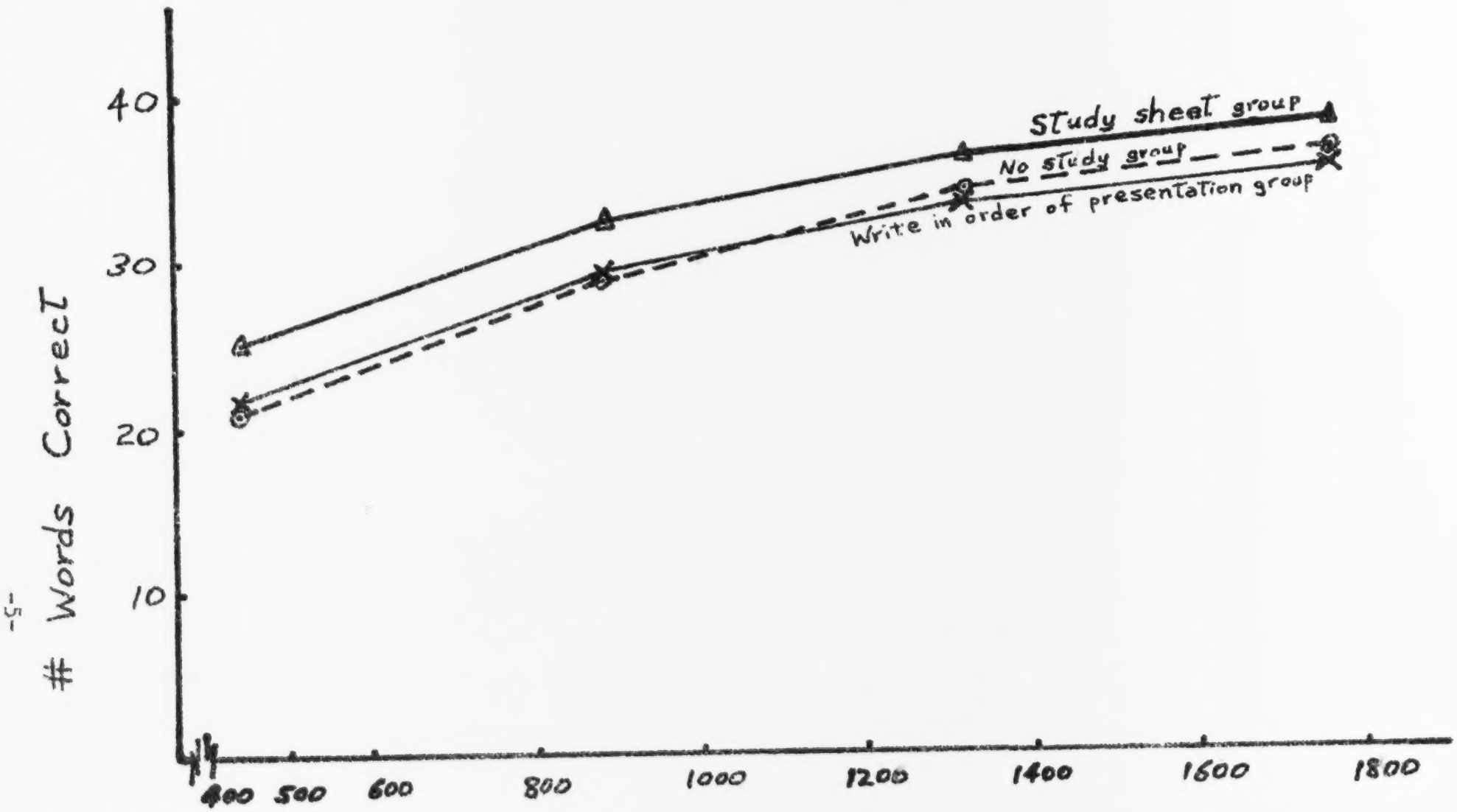


Fig. 1. Number of words correct as a function of time in the learning situation for the three different groups

remaining thirteen Ss organized on the basis of subjective meaning (sub-group MO). For MO Ss the categories were identified, in the post-experimental session, by such names as "food items, utencils, animals, sorority, structures, miscellaneous, made-a-story," etc. Collapsing across the four trials, an analysis of variance indicates significant differences amongst the total number of words correct for the three sub-groups ($F = 6.76$, $df = 2,21$, $p < 0.01$), and the Newman-Keuls procedure indicates significant ($p < 0.01$) differences between NO and AO, and between NO and MO, but no significant ($p > 0.05$) difference between MO and AO. The means for total number of words correct are: AO = 137.6 MO = 133.8, and NO = 119.5. The six Ss who showed no organization on their study sheets (sub-group NO) averaged fewer correct on each of the four trials, attaining an average of about 36.5 correct on trial four. Their learning curve shows very little negative acceleration across the four trials. By contrast the learning curves for the AO and MO sub-groups show sharp negative acceleration, and an obvious ceiling effect as they average about 39 (out of a possible 40) correct by trial four. The results are depicted in Figure 2.

- - - - - - - - - - - - - - -
Insert Figure 2 about here
- - - - - - - - - - - - - - -

The relations between proportion of items correct and the characteristics of the E-defined categories were examined separately for the NO and MO sub-groups of the Experimental group. The data for the alphabetizers (AO) was not examined in this analysis since their mode of subjective organization specifically disrupts the E-defined categories. Figure 3 depicts the results for the four different kinds of E-defined categories "built-in" to the list of forty words. The ordinate is the proportion of items correct (collapsed across all four trials) for each kind of category,

- - - - - - - - - - - - - - -
Insert Figure 3 about here
- - - - - - - - - - - - - - -

for each of the two sub-groups. The difference between high and low dominance categories for the 13 Ss of the MO sub-group is significant ($t = 3.09$, $df = 12$, $p < 0.01$), as is the same comparison for the 6 Ss of the NO sub-group ($t = 2.95$, $df = 5$, $0.02 < p < 0.05$). There are no differences (to the second decimal place) in proportion correct between words in low dominance categories and words unrelated according to the dominance norms. The words in the two "cue categories" (i.e., table utensils and class days) were correct for almost all of the Ss of both sub-groups from trial one on. Of the six Ss in the NO sub-group one S missed one of the six words on trial two. Of the thirteen Ss in the MO sub-group two missed the three table utensil words on trial two, while a third S missed them on trial three. Thus, the words of the "cue categories" showed essentially perfect learning from trial one through trial four. The words of the other categories showed "typical" learning curves across the four trials.

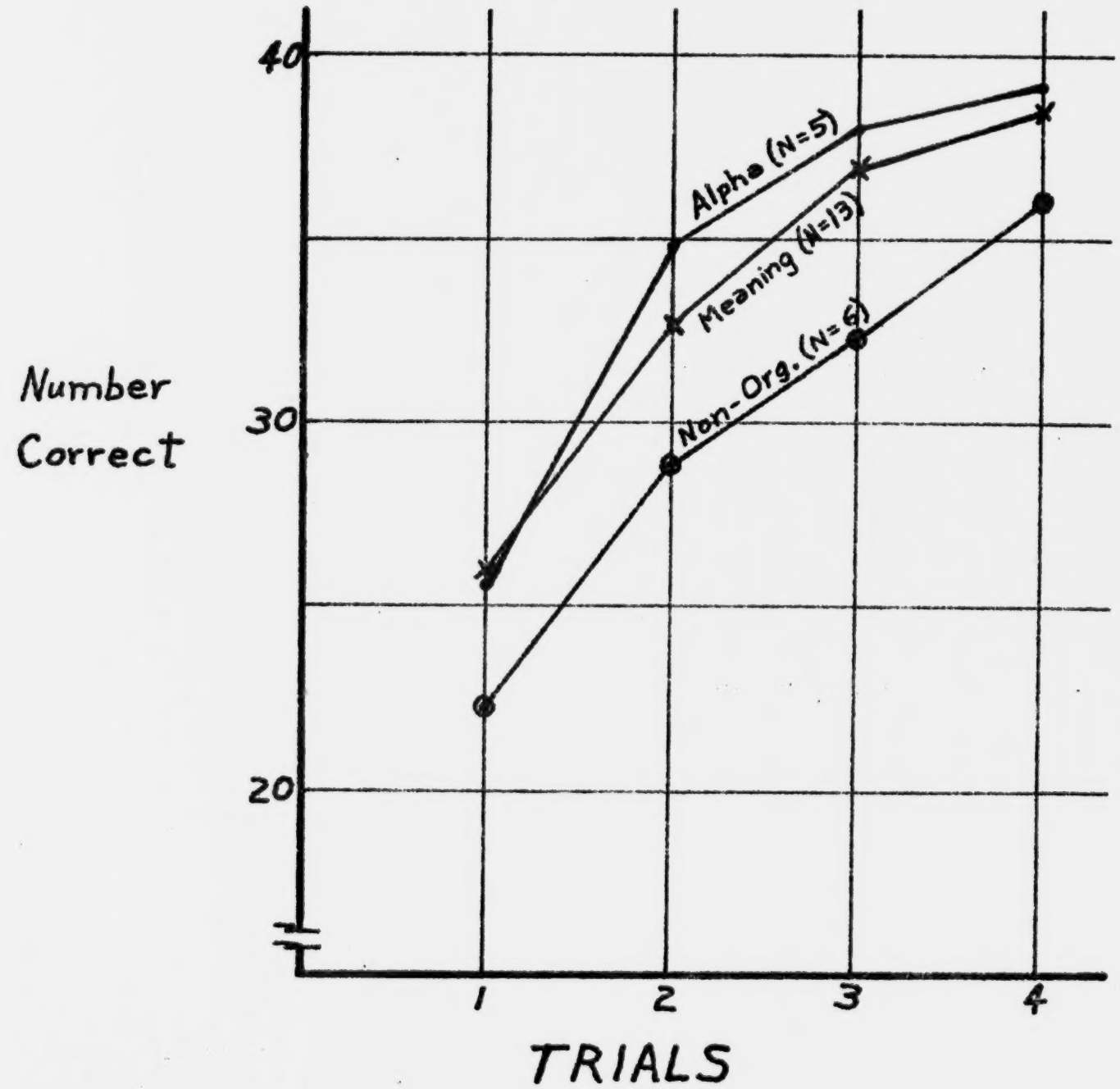


Fig. 2. Number Correct as a function of trials for the three sub-groups in the study-sheet group. Sub-groups defined in terms of organization displayed on the study sheets.

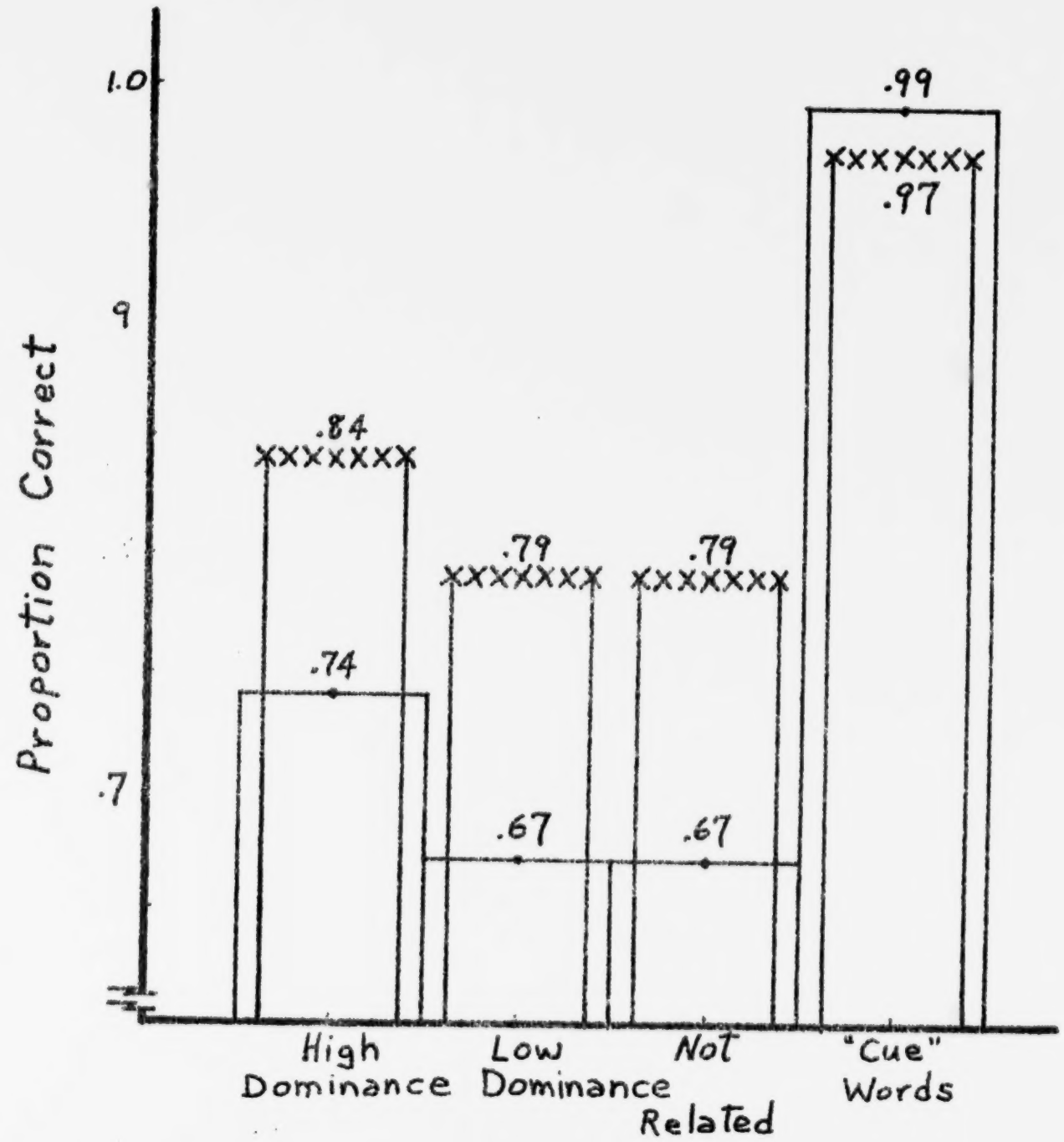


Fig. 3 Proportion of items correct as a function of the character of the experimenter-defined category of which the item was a member. Ss who showed subjective organization on their study sheets (xxx) are contrasted with those who did not.

The degree to which words cluster on the tests may be examined by counting the number of times words in the same category are adjacent to each other in the test lists. This count of number of adjacencies is referred to as repetitions observed (RO). If the words on a test are rearranged so that all words in the same categories are adjacent to each other, and then a count made of the number adjacencies, the measure referred to as repetitions possible (RP) is obtained. RP is directly related to number of correct words on the test. It may be calculated by the formula $RP = \sum_{i=1}^{k} (m_i - 1)$, where m_i is the number of words in the i^{th} category that appeared on the test and k is the number of different categories represented by at least one word on the test. Fig. 4 depicts

- - - - - - - - - - - - - - -
Insert Figure 4 about here
- - - - - - - - - - - - - - -

one way of examining the relationships between RO, RP, and the other variables of the experiment. The RO variable is on the ordinate, the RP variable on the abscissa, and the 45 degree line represents the locus of points defined by perfect, or total, clustering (RO = RP).

Fig. 4a depicts trial-two data for the 13 Ss of the MO sub-group. The circles indicate clustering scores based on the E-defined categories built into the list. The Xs indicate clustering scores based on the Ss own categories as defined on his study for that trial. Fig. 4b depicts the same things for trial four. When categories are defined by the Ss the cluster of points moves closer to the RO = RP line as learning progresses. When the categories are defined by E the cluster of points moves further away from the RO = RP line as learning progresses. Fig. 4c depicts this shifting in terms of group averages for the four trials. Once again the circles indicate clustering scores based on the E-defined categories and the Xs indicate clustering scores based on the study-sheet-defined categories. The steady progression over trials towards perfect clustering for the study-sheet-defined categories, and the steady progression over trials away from perfect clustering for the E-defined categories make statistical analysis appear superfluous. The distance of each point from the RO = RP line (along a perpendicular to the RO = RP line) may be shown to be equal to $(1/\sqrt{2}) \cdot (RP - RO)$. An analysis of variance of the (RP - RO) measures supports the obvious in Fig. 4c. The trials-by-definitions interaction is significant ($F = 28.8$, $df = 3,32$*, $p < 0.001$). For the E-defined categories the mean (RP - RO)'s for the four trials are 9.0, 15.1, 17.2, and 17.9, and the trials effect is significant ($F = 46.5$, $df = 3,36$, $p < 0.001$). For the study-sheet defined

*With 13 Ss and four trials, the df should be 3,36. RP - RO values were not available for four Ss on trial one (they wrote the words in the arbitrary order in which they were presented). Average trial-one values were used for these missing scores in the analysis, and four df subtracted in computing the error mean-squares.

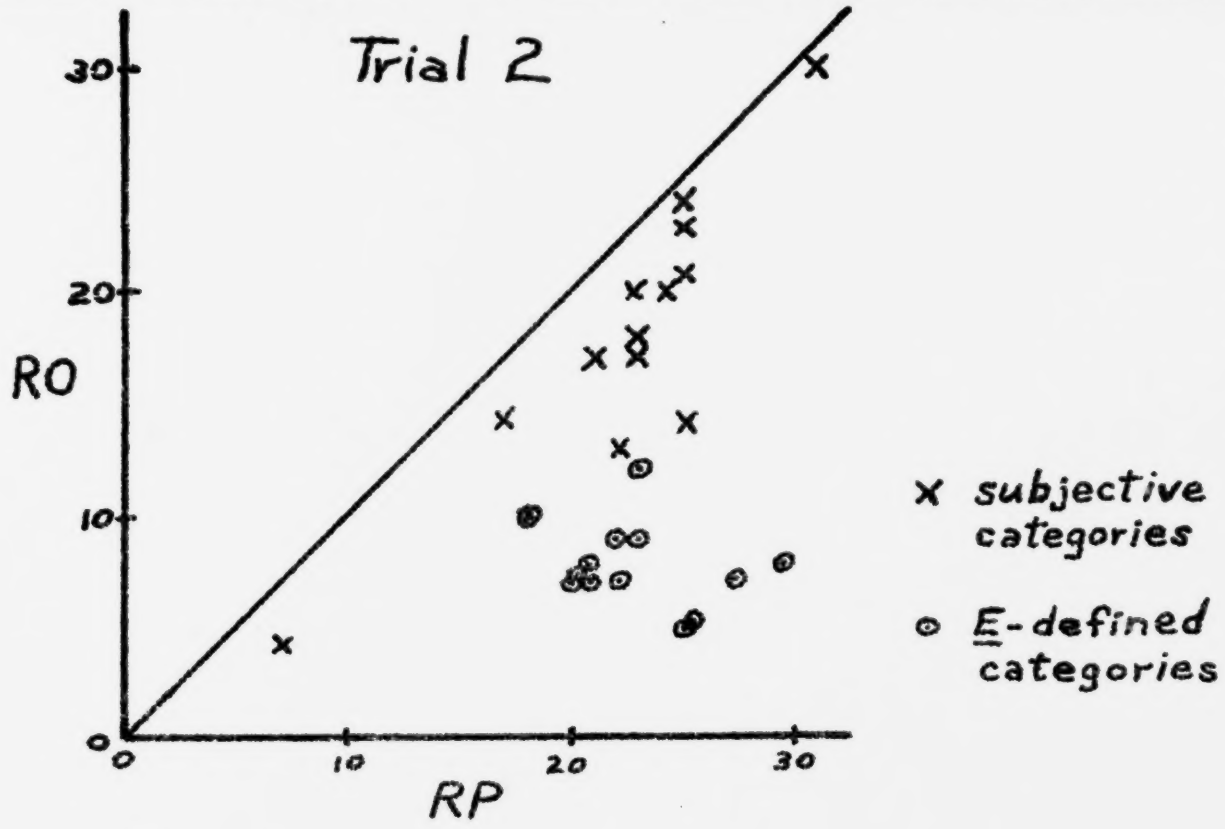


Fig. 4 a.

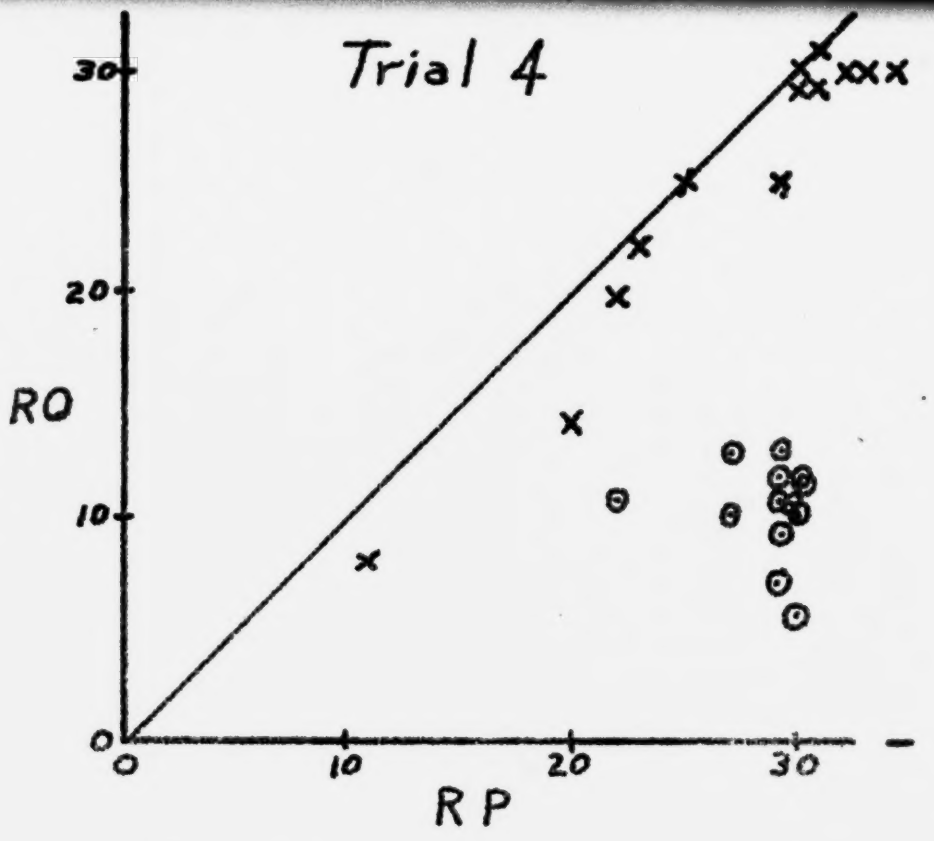


Fig. 4 b.

Fig. 4. Observed cluster score (RO) relative to possible cluster score (RP). Clusters defined subjectively (via study sheets) or defined by E.

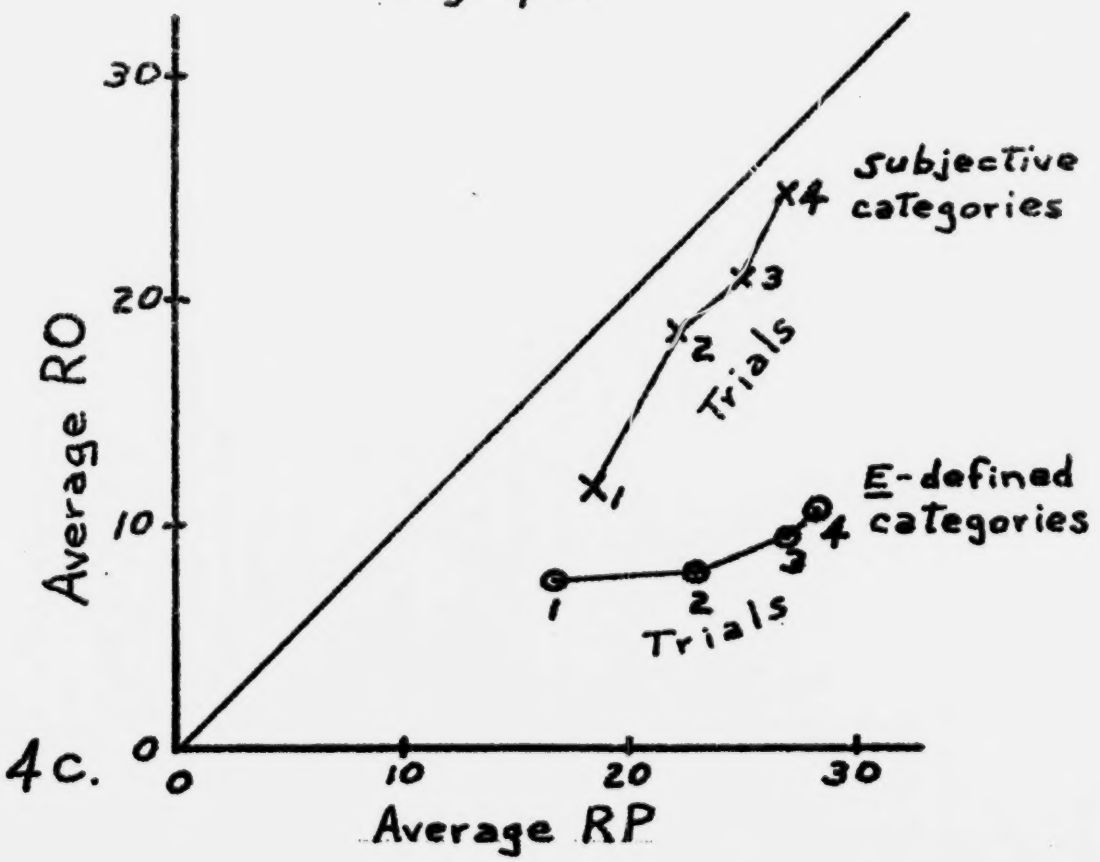


Fig. 4 c.

categories the corresponding means are 6.2, 4.3, 4.5, and 2.2, and the trials effect is also significant ($F = 4.62$, $df = 3,32^*$, $0.005 < p < 0.01$).

Degree of clustering as a function of the character of the E-defined category of which the items were members is depicted in Fig. 5, in a manner analogous to that in Fig. 3 for proportion of items correct.

- - - - - - - - - - - - - - - -
Insert Figure 5 about here
- - - - - - - - - - - - - - - -

These are the data for sub-groups MO and NO. The value on the ordinate of Fig. 5 is defined as follows: for each S, for each E-defined category of words, the (RP - RO) measure is summed across the four trials and the sum is then divided by the sum of the corresponding RP measures. These proportions are the basic data for this analysis. They represent the distance from perfect clustering, collapsed across trials, relative to the maximum distance possible given the Ss particular performance on his four tests. In formula the ratio may be represented as:

$$\frac{\overbrace{\Sigma(RP - RO)}^{4\ \text{trials}}}{\underbrace{\Sigma\ RP}_{4\ \text{trials}}} = 1 - \frac{\overbrace{\Sigma\ RO}^{4\ \text{trials}}}{\underbrace{\Sigma\ RP}_{4\ \text{trials}}}$$

for each S, for each E-defined category of words. The values plotted in Fig. 5 are simply averages of these ratios across appropriate Ss. The ordinate scale is inverted so that "more clustering" is "higher" on the ordinate. There are no significant differences in clustering amongst the dominance-defined categories of words (not even one of the possible within-S t-tests had a $p < 0.10$), despite the fact that high and low dominance words did differ in terms of proportion correct (see Fig. 3 and associated analysis). These results are in essential agreement with those of Bousfield and Puff (1964). The "cue" words, i.e., table utensils and class days, show essentially perfect clustering.

The two control groups and the experimental sub-group NO may be examined ror total number correct (collapsed across trials) and for total (RP - RO) scores for the E-defined categories. There are no obvious differences amongst the three groups on either measure, but all three are different than the experimental sub-group MO on both measures. For the number correct measure the means are: 119.25 for Control 1 ($n = 24$), 119.50 for sub-group NO ($n = 6$), 119.89 for Control 2 ($n = 21$), and 133.85 for sub-group MO ($n = 13$). Differences in variances are not significant ($0.05 < p < 0.10$ via Bartlett's test), and an analysis of variance yields an $F = 6.46$ ($df = 3,60$) for between groups ($p < 0.01$). The Newman-Keuls procedure indicates the MO mean different from the other three ($p < 0.01$),

*See page 9 for footnote

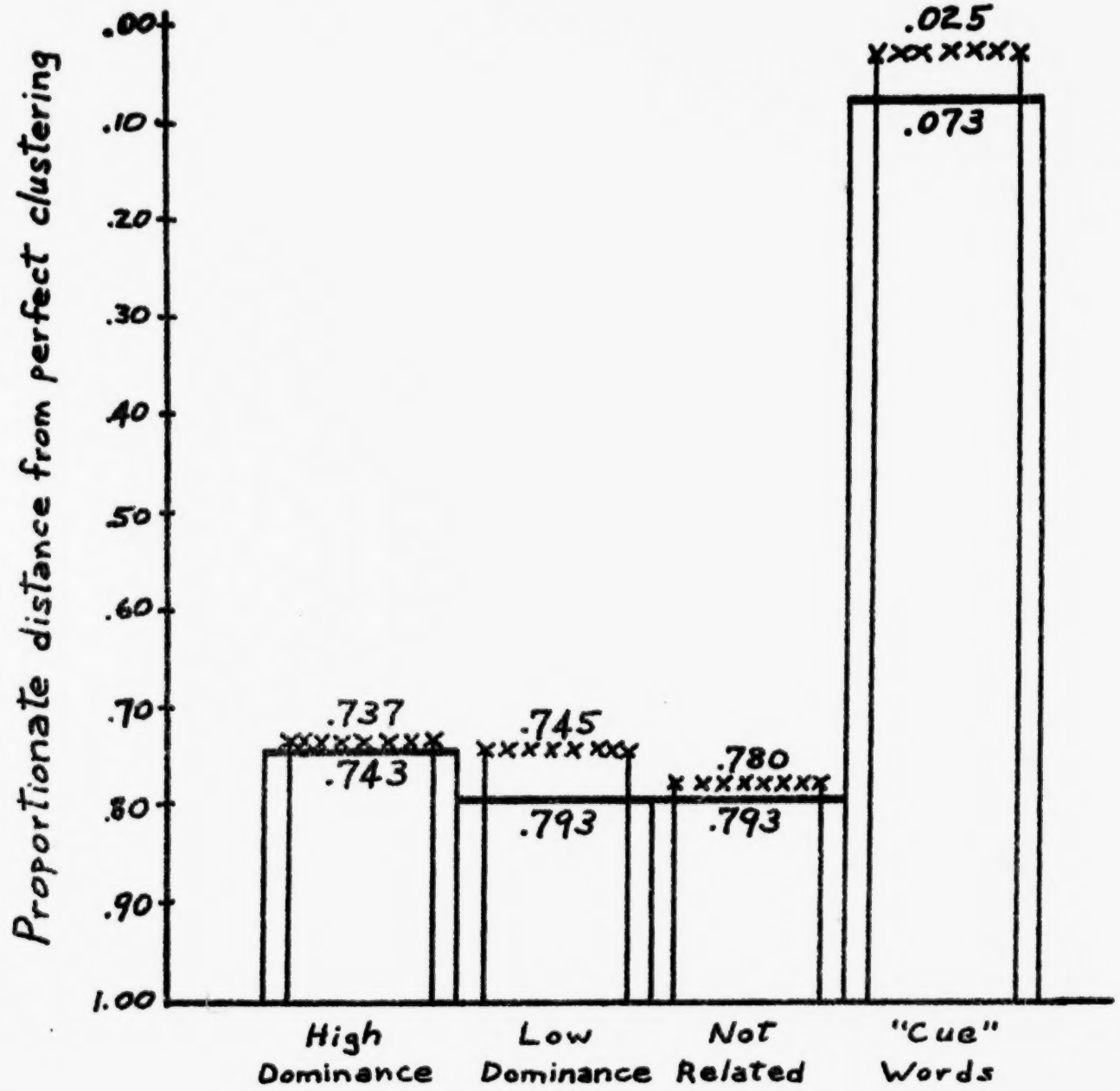


Fig. 5. Degree of clustering of items as a function of the character of the E-defined category of which they were members. Ss who showed subjective organization on their study sheets (xxx) are contrasted with those who did not.

but no differences amongst those three ($p > 0.05$). For the (RP - RO) measure the means are: 49.00 for sub-group NO ($n = 6$), 50.43 for Control 2 ($n = 21$), 51.75 for Control 1 ($n = 24$), and 59.23 for sub-group MO ($n = 13$). Differences in variances are not significant ($p > 0.05$ by F_{max} test), and an analysis of variance yields an $F = 3.03$ ($df = 3,60$) for between groups ($0.01 < p < 0.05$). The Newman-Keuls procedure indicates the MO mean significantly different from the NO mean ($p < 0.05$), differences between the MO mean and the two control group means approach significance (p just slightly greater than 0.05), and differences amongst the NO, Control 1, and Control 2 means not significant (p much greater than 0.05).

Discussion for Experiment One

The opportunity to overtly organize (on study sheets) the material to be learned facilitated the learning of the material. Simply writing the words in their random orders of presentation did not facilitate learning relative to the standard free-recall (no study writing at all) conditions. Further, only those Ss who took advantage of the opportunity to overtly organize the material (sub-groups MO and AO) were the ones to show more rapid acquisition of the material. Those Ss who were given the oopportunity to overtly organize the material, but who failed to utilize this opportunity (sub-group NO), showed acquisition performance indistinguishable from those Ss not given the opportunity to overtly organize (Control 1 and Control 2). The same general relationships amongst these groups and sub-groups are also true in terms of the clustering behavior (for the E defined categories).

Though the number of Ss who utilized alphabetic organization ($n = 5$) was smaller than the number who utilized organization according to subjective meaning ($n = 13$), the present experiment indicates no significant difference in learning performance for the two modes of organization.

The E-defined categories based on the level of dominance definitions were, in general, not utilized by the Ss. Though the high dominance words were somewhat easier to learn than the low dominance and non-related (according to dominance) words, there was no difference in clustering for these different levels of dominance. Further, clustering performance in terms of these "built-in" categories actually indicates a decrease in the utilization of these categories as learning progresses. The category definitions based on dominance are not "persuasive." Ss tend to ignore them, and appear to find other criteria for categorizing or organizing the words. Bousfield and Puff (1964) report contrasting results.

In contrast to the dominance-defined categories the two "cue categories" were highly salient and "persuasive." Almost all Ss showed perfect retention for these words from trial one on, and sub-group MO also showed essentially perfect clustering from trial one on. Sub-group NO and the two Control groups show a rapid increase in clustering for these words, reaching essentially perfect clustering by trial two or three (this last observation is just descriptive, i.e., it is not analyzed statistically and is, therefore, not reported in the results section).

Thus, it is possible for E to define categories which most Ss will adopt in their subjective organization of material to be learned.

The marked difference in subjectively defined clustering relative to E-defined clustering leads to the obvious conclusion that Ss may be ignoring the E-defined independent variable. The present paradigm provides for a check on the extent to which this may be true, and for the development of E-defined variables of varying and known degrees of "persuasiveness."

It was originally thought that the study-sheet paradigm would provide a "sneak look inside the S," if you will, in the usual free-recall experiment. This is obviously not so. The opportunity to organize on their study sheets changes the S's behavior. For example, (1) no control S alphabetized, but five experimental Ss did, and (2) experimental Ss got more words correct. It is proposed, however, that the experimental paradigm presented is no less interesting than the standard free-recall situation, or any other standard learning paradigm for that matter.

Experiment Two

Since the concept dominance variable was impotent with respect to clustering or "organizing" behavior another E-defined variable was sought. Further, generalization of the findings of experiment one required at least one other set of stimulus materials. The conclusion that the study-sheet paradigm was not providing a "look inside of" the standard free recall paradigm led to considerations for maximizing the usefulness of the overt organizing behavior of the Ss. Performance on study sheet one of the first experiment was essentially useless since Ss didn't know the total composition of the list to be memorized until after they had completed that first study sheet. To eliminate this problem all Ss were given a "pre-look" at the total list of words to be memorized. This was done prior to the first trial only, and the words were randomly arranged so as to continue not suggesting any particular organization. Instructions remained "arrange the words on the study sheet to best help you memorize." The list was made longer in order to avoid the ceiling effect exhibited by most of the experimental Ss of experiment one by trials three and four. Control groups were eliminated and all Ss run under study-sheet conditions in order to maximize the quantity of study-sheet data on which Ss "organized according to meaning." This larger number of completed study sheets was used to develop a more complete objective procedure for establishing the subjective categories of each S.

Method for Experiment Two. Fifty-eight Ss yielded the data of experiment two, all of them learning via the study-sheet paradigm. One subgroup of 21 Ss had four trials. Two subgroups of 15 and 22 Ss each had five trials.

The materials of experiment two consisted of 72 words chosen from those utilized by Marshall (1967) in the study previously referred to. The 72 words were really 36 pairs of words; the 36 pairs being divided into subsets of six pairs each. Each subset of six pairs differed from the next in terms of the range of the Mutual Relatedness (MR) index

between the words of the pairs. Thus the pair "man-woman" was in the subset of six pairs having a high degree of MR, while the pair "minute-day" for example was at the opposite end of the continuum. The MR index is based on normative association data and reflects the degree to which each word of the pair elicits the other, and words in common, relative to all words elicited by both members of the pair.

Further, each subset of six pairs was divided into two sub-subsets of three pairs each, one of the sub-subsets involving pairs which might be called categorized, that is, each member of the pair could easily be incorporated within one category name, e.g., man and woman are both human beings. These are contrasted with the non-categorized pairs, e.g., spider-web or food-eat. The 36 pairs of words are listed in Table 2, arranged according to MR level, and separated into categorized and non-categorized groups. In the Marshall study previously referred to a different group of

- - - - - - - - - - - - - -
Insert Table 2 about here
- - - - - - - - - - - - - -

Ss was utilized for each MR level, with twice as many pairs of words at each level than was used in the present experiment. In the present experiment, however, all Ss were presented with pairs of words at all MR levels, thus yielding within S comparisons across MR levels, whereas the Marshall study yielded between S comparisons across MR levels. Further, the Marshall study utilized the relatively standard free recall learning paradigm while the present study employed both a "pre-look" and the study-sheet paradigm.

There was 1 1/2 min. for the pre-look. The presentation was at the rate of 5 sec. per word (six min. to present all 72 words once). There was 1 1/2 min. for study of the study sheet, and the test was timed for five min. Repeats of all but the pre-look made up the subsequent trials. Several different random arrangements of the words on a blank study sheet were used for the pre-look. All Ss saw the same series of random sequences of words during the presentation periods of the successive trials.

All pre-look, study, and test sheets were on 8 1/2" x 14" ("legal" size) sheets. The study sheets had 12 columns each about 1 1/16" wide, and 28 rows each about 1/4" high, thus outlining 12 x 28 cells each about 1 1/16" x 1/4". For the pre-look sheet the 72 words were randomly assigned to 72 out of the 336 cells. The test sheets contained two long columns of numbered spaces, 1 - 36 and 37 - 72.

The procedure for objectively establishing the subjective categories from the individually prepared study sheets involved the following: the first step was to search all study sheets to determine whether or not the words were written down in the order in which they were presented. If 90% or more of the words were written in the order in which they were presented, either by column or by row, then the study sheet was classified as an "order of presentation" study sheet, and for that S on that trial there was no information for determining subjective categories. Twenty-three

TABLE II. The 36 pairs of words used in Experiment Two, classified according to Mutual Relatedness (MR) level and whether or not the pair was a member of an obvious category.

		Mutual Relatedness Level					
		Low					High
	CATEGORIZED	(0.0 - 4.9)	(5.0 - 9.9)	(10.0 - 19.9)	(20.0 - 29.9)	(30.0 - 39.9)	≥ 40.0
		minute	trumpet	hammer	eagle	doctor	arm
		day	ukelele	pliers	crow	nurse	leg
		toe	bee	tin	baby	violin	man
		head	fly	iron	child	piano	woman
		scissors	Mexico	dress	foot	daisy	thought
		needle	India	sweater	knee	gardenia	idea
NON	CATEGORIZED	image	bumpy	high	blossom	blow	picture
		wall	silk	mountain	vase	horn	frame
		train	cottage	fish	drive	dream	eat
		tweet	cheese	boat	car	sleep	food
		warm	paper	glass	cloak	apple	spider
		glove	bag	milk	dagger	cider	web

subject-trials out of a total of 269 fell into this category. For the remaining 246 subject-trials the determination of which words the S intended to group together was made exclusively in terms of the geometry of filled and empty cells on the study sheets, that is, the particular contents of the cells was completely ignored. Cells adjacent to each other were scored as "belonging together." The words in them, therefore, were scored as members of a single subjective category. The problem was one of determining whether it was to be horizontal or vertical adjacency that would be used in the scoring. Five pairs of measures were used for the determination. The first involved a count of the number of adjacent filled cells in going down each column, and then an equivalent count in going across each row. The second involved a count of the number of transitions from a filled cell to an empty cell as the study sheet was examined, first column by column, and then row by row. The third measure involved a count of the number of words in the upper-most row of the study sheet which was utilized by the S, thus defining a margin count for column organization. A similar count was made of the number of words in the left-most column utilized, thus finding a margin count for row organization. The fourth measure involved examining for isolated groups of filled cells, that is, filled cells surrounded by spaces and/or margins. A separate count was then made of the filled cells in these isolated clusters, first by column and then by row. The fifth and last pair of measures involved the variance of the counts going into the first pair of measures. The next step involved inserting these five pairs of numbers into a somewhat complex logic program, and the outcome was the classification of the S into either a row organizer or a column organizer for that study sheet. After eliminating two Ss out of the original sixty for failing to follow instructions it was possible to build the logic ad hoc so as to successfully classify all 269 subject-trials involved. The cross validation on an additional sample remains to be done, but success in 269 out of 269 cases is quite promising. Success, of course, is here defined in terms of agreement with Es' judgments based on examining the study sheet and the contents of the cells. Incidentally, most Ss were column organizers and very, very few had any mixed modes of organization.

Results for Experiment Two

Despite large procedural differences between the Marshall (1967) study and the present one, the results are remarkably similar in terms of the E-defined variables. Fig. 6 depicts the average number of words correct

- - - - - - - - - - - - - - -
Insert Figure 6 about here
- - - - - - - - - - - - - - -

for the categorized versus the non-categorized words in the list. In both studies categorized words were recalled better than non-categorized words. But the most interesting finding common to both studies is depicted in Fig. 7. In this figure there is plotted, for each of the five trials, an

- - - - - - - - - - - - - - -
Insert Figure 7 about here
- - - - - - - - - - - - - - -

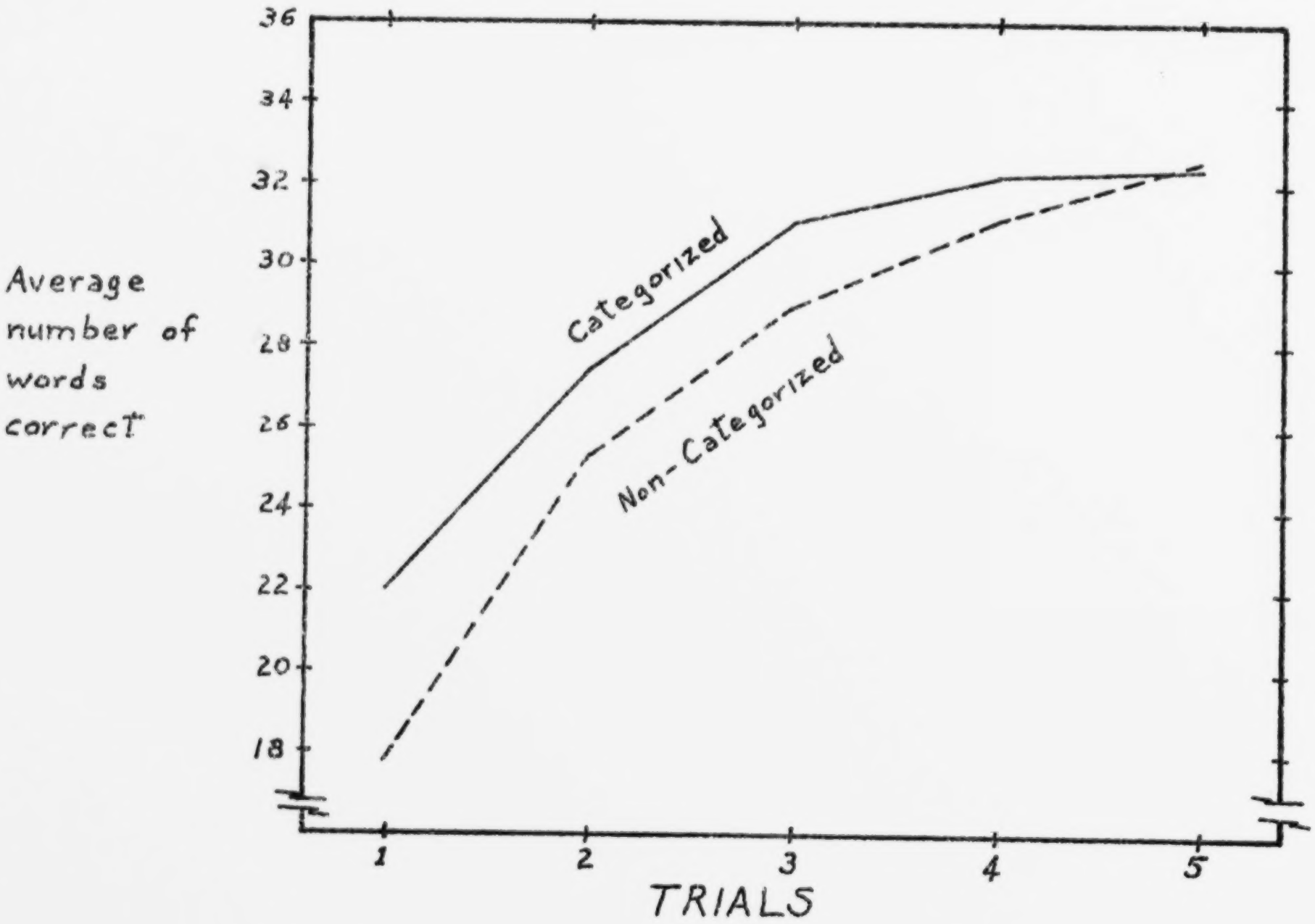


Fig. 6. The average number of words correct as a function of trials for categorized and non-categorized words. N=59 for trials 1 through 4, and N=37 for trial 5.

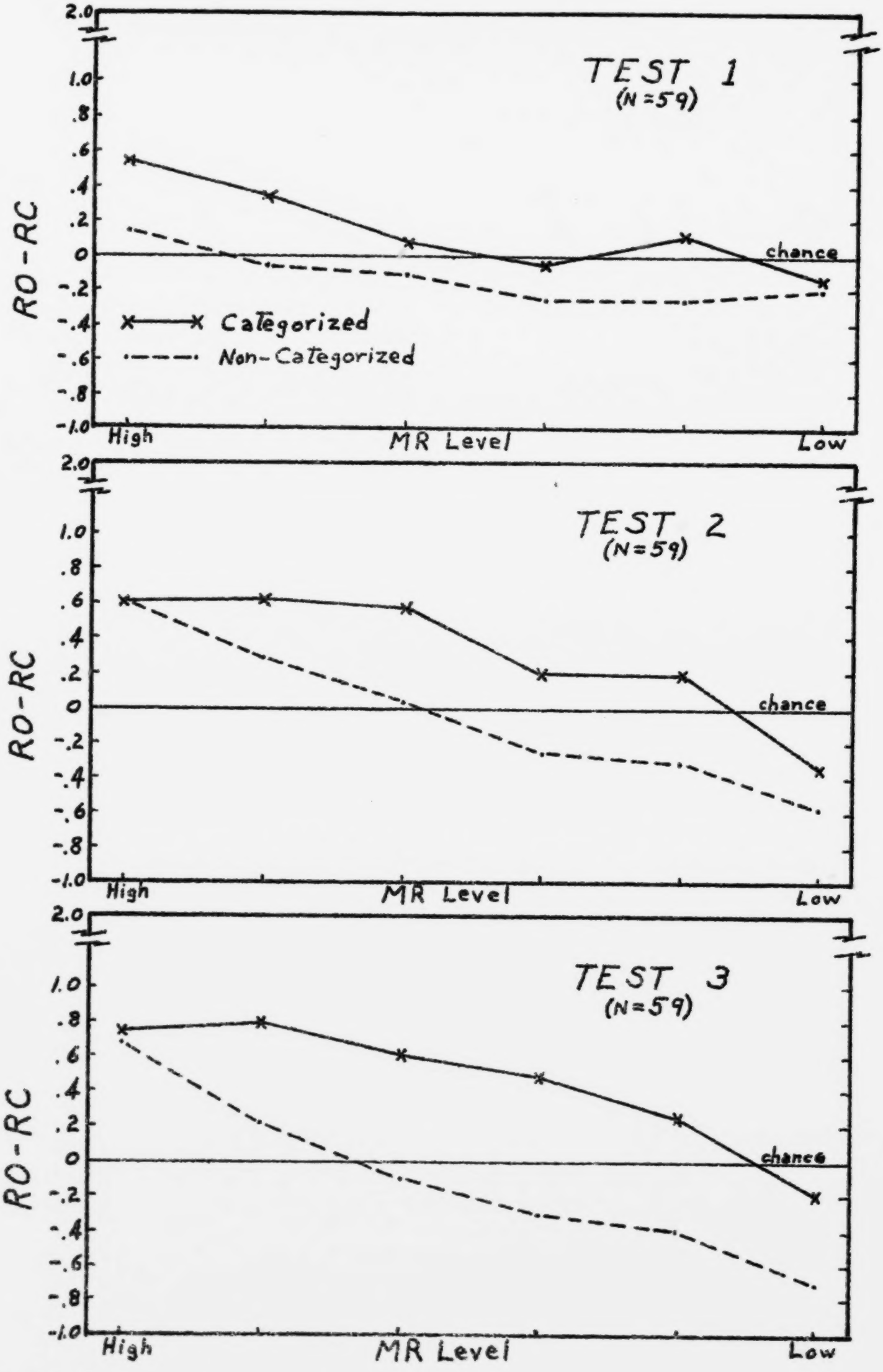


Fig. 7 (continued on next page)

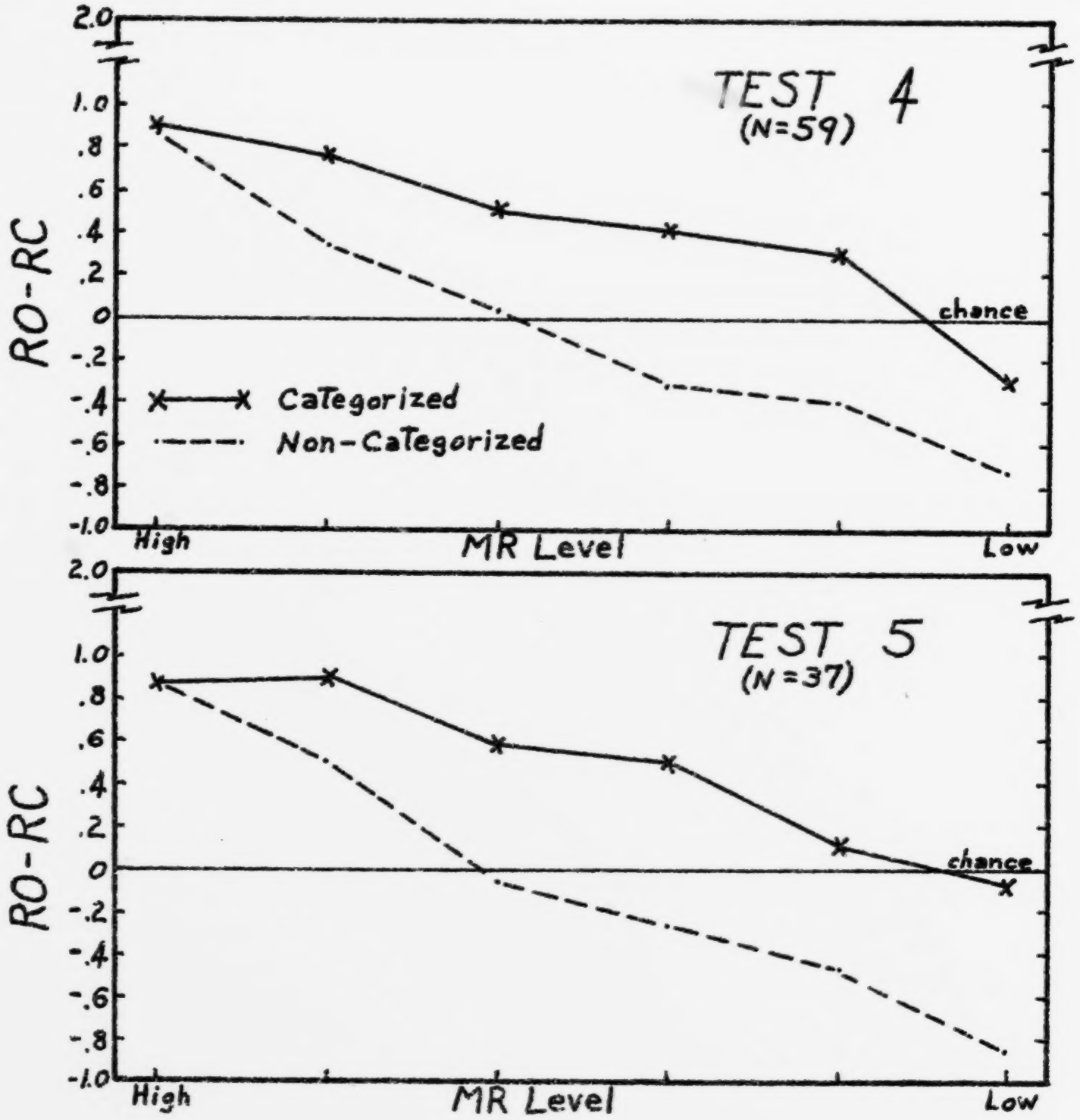


Fig. 7. (part 2 of 2). Experimenter defined test-sheet organization as a function of Mutual Relatedness (MR) Level. Pairs of Categorized and Non-Categorized words are equated for MR Level. The index for organization is the number of repetitions observed (RC) minus the number of repetitions expected on the basis of chance (RC). For the three pairs of words at each point the maximum value of the index is 2.0, the minimum value is -1.0, and the chance value is 0.0.

index of clustering, on the ordinate, as a function of MR level, on the abscissa; with the parameter within each graph being the distinction between categorized and non-categorized pairs. The index of clustering is the average number of repetitions obserted (RO), minus that number of repetitions (RC) which would be expected on the basis of chance alone given the particular set of words the S produced on his test. This adjustment for chance is based on Bousfield's formula as reported by Dallet (1964). The formula is:

$$RC = \frac{\sum_i (n_i)^2}{\sum_i n_i} - 1,$$

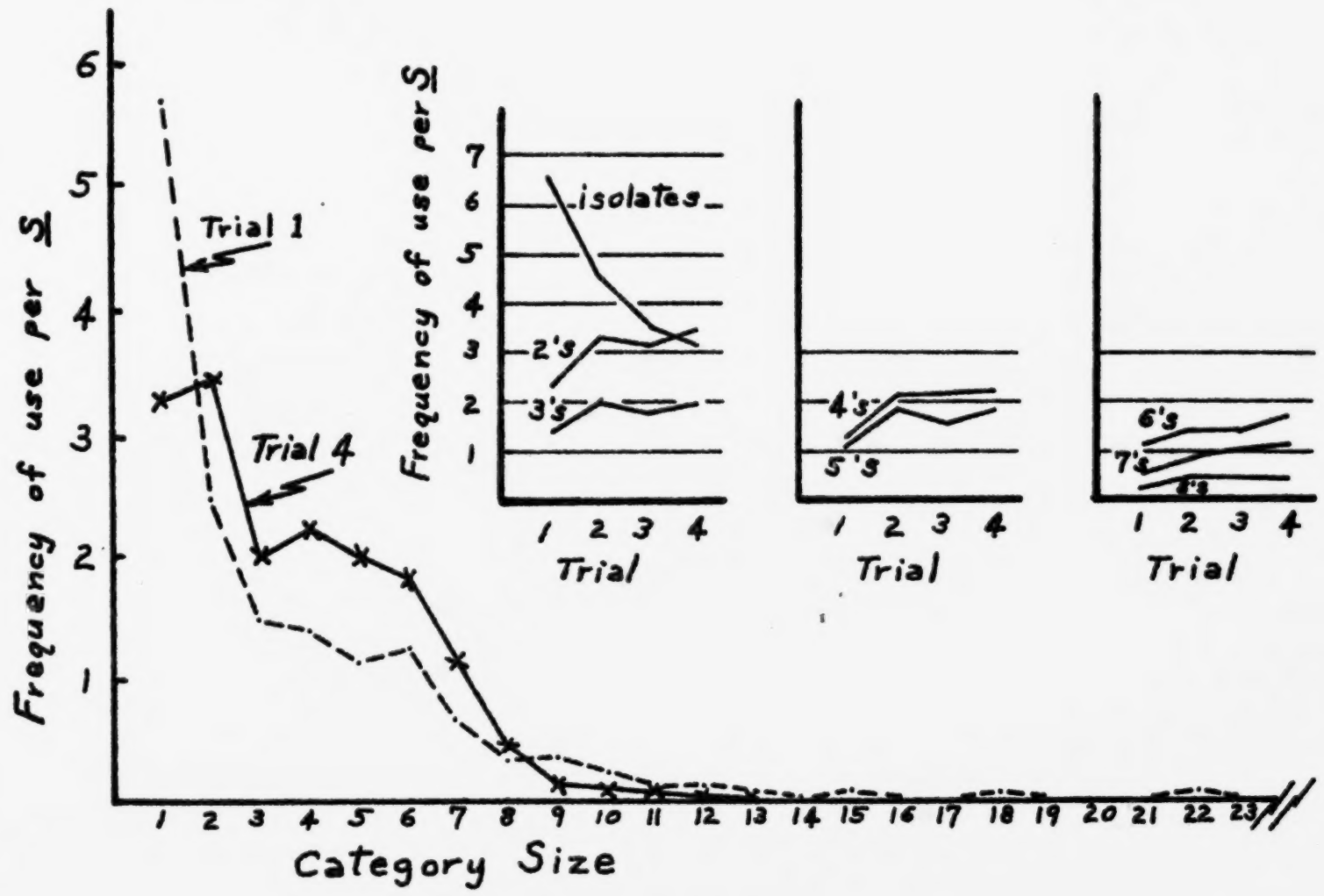
where n_i is the number of words in the i^{th} category which appeared on the test. The derivation and extension to the present case is included in Appendix B. The relationships depicted in Fig. 7 may be summarized as follows: beyong trial one the figures consistently indicate little or no distinction between categorized and non-categorized pairs at the highest MR level, with a widening distinction between the two kinds of pairs as MR level decreases to the low end. These results agree very nicely with those reported by Cofer (1965) and Marshall (1967), despite the fact that a different index of clustering was used in that experiment, and despite all of the procedural differences involved in the two experiments. However, an additional note to keep in mind with respect to Fig. 7 is the movement of the lines from test to test relative to the zero or chance line, and relative to the 2.0 or maximum possible upper limit. In general, the low MR pairs, particularly the non-categorized ones, cluster less than would be expected by chance, and this negative cluster score actually increases in magnitude from test to test. It is somewhat offset by corresponding increases relative to chance for the words at the higher end of the MR scale, in particular the categorized words. However, note that nowhere is a point to be found above 1.0 on the dependent measure scale, with a value of 2.0 being the score that would be obtained if perfect clustering occurred. As a matter of fact, if the data are examined in terms of a slightly different measure, namely the one used in Experiment One, i.e., the number of repetitions possible (RP) minus the number of repetitions observed (RO), one finds that as learning progresses the difference actually grows. Thus, collapsing over the MR levels, the overall E-defined clustering actually decreases with learning. This is not true for subjectively defined clustering. The results are very similar to those for Experiment One as depicted in Fig. 4c.

Fig. 8 depicts some of the characteristics of the subjective clustering as measured on the study sheets. The large upper plot simply

- - - - - - - - - - - - - - -

Insert Figure 8 about here

- - - - - - - - - - - - - - -



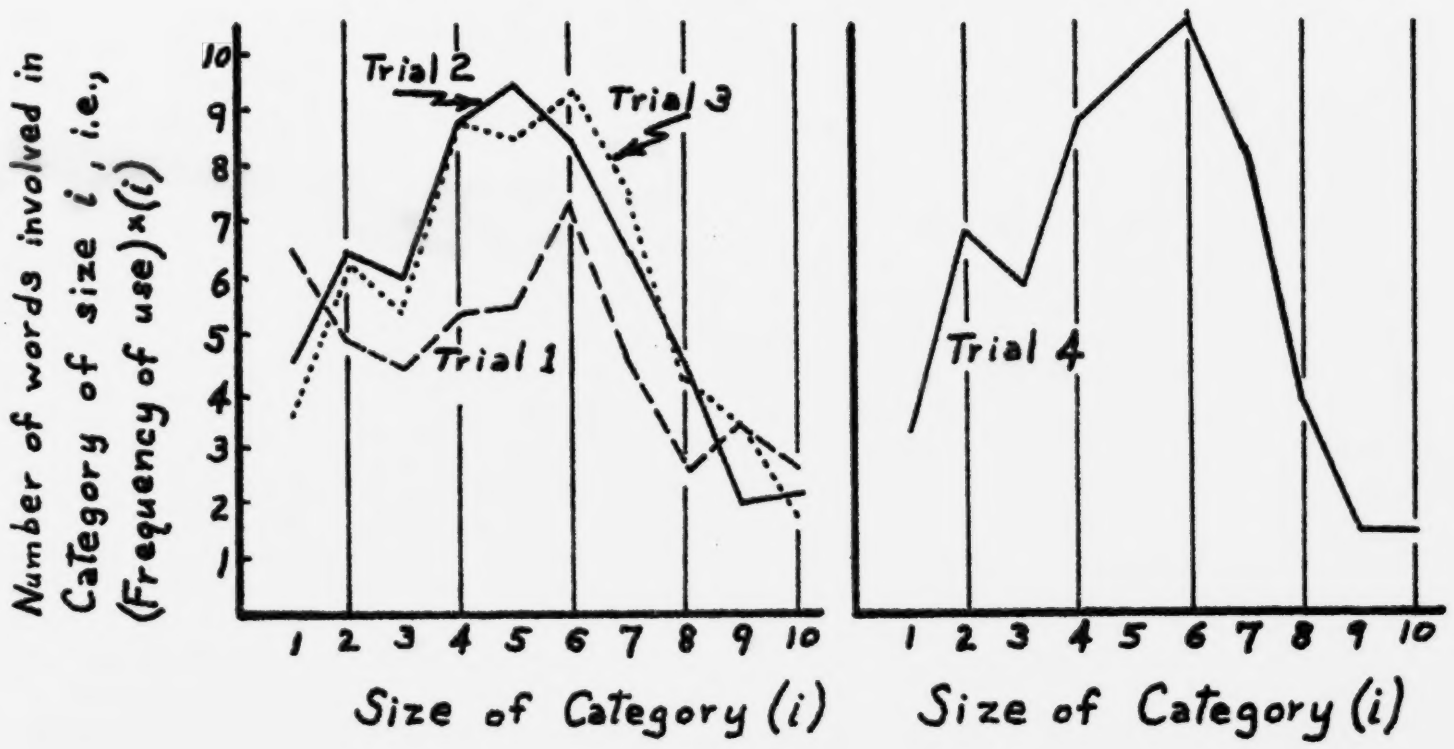

Number of words involved in Category of size i, i.e., (Frequency of use) x (i)
Trial 2
Trial 3
Trial 1
Trial 4
Size of Category (i)
Size of Category (i)

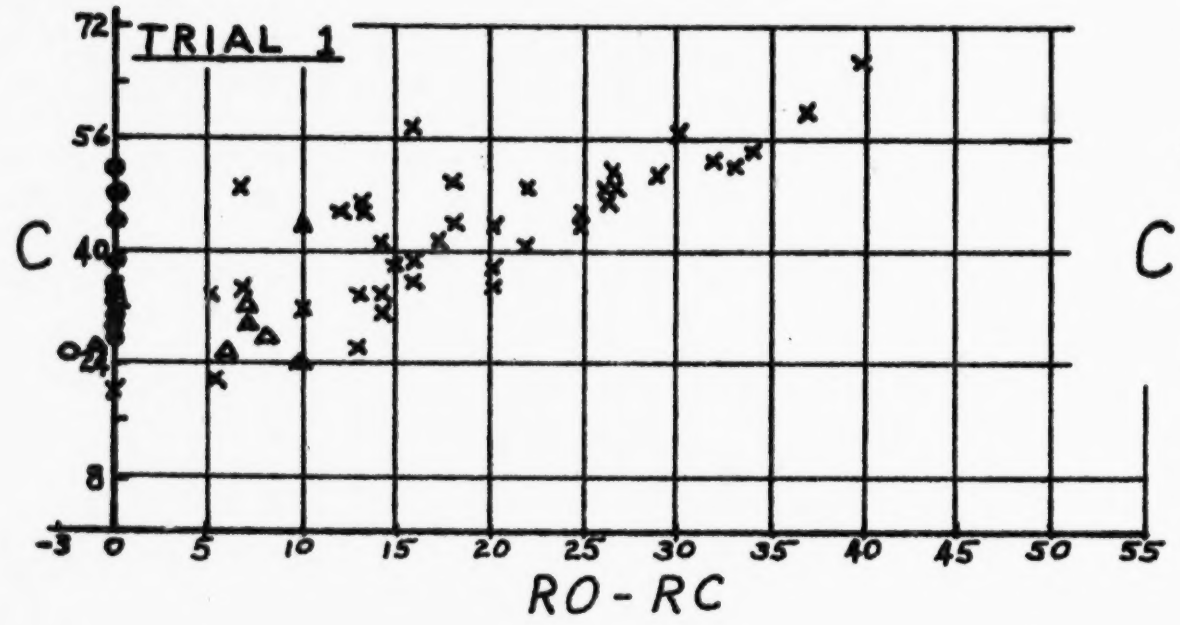
Fig. 8. Frequency of use and number of words involved as a function of category size (subjective categories defined via study-sheet performance). Data are shown for trials one through four.

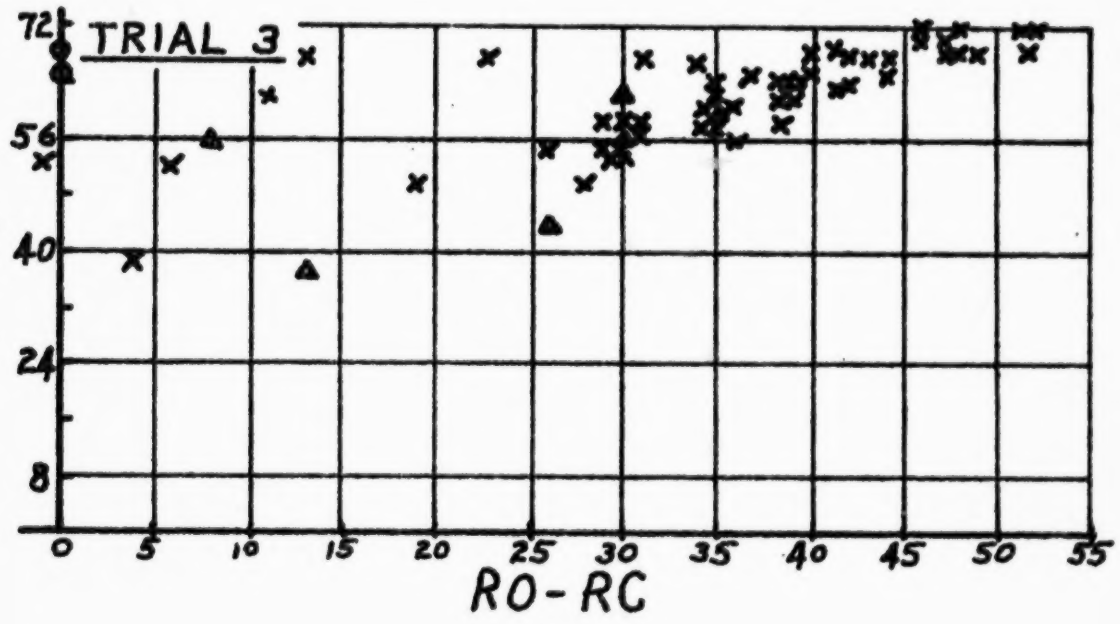
gives the average frequency of use of subjective categories of various sizes, for trial one and for trial four. The three smaller insert plots trace the average frequency of use across four trials, and do so for isolated words and for categories of two through eight words. The number of isolated words on the study sheets decreases as a function of trials. All subjective categories of two through eight words show increased average frequency of usage as a function of trials. The dip in frequency of usage of categories of three words is believed to be due to the pairs of words "built-in" to the stimulus list. The lower two plots of Fig. 8 have as their dependent measure the average number of words involved in the categories of various sizes, as specified along the abscissa. The dependent measure is simply the average frequency of usage (of the upper plots) times the size of the category. (Once again the dip for categories of size three is believed to be due to the built-in pairs of words. A similar plot for the data of Experiment One, in which the minimum built-in category had three words, shows a dip for categories of two words.) The results for four trials are depicted in the two plots. The means of the distributions move toward larger categories as a function of trials. By trial four approximately half of the 72 words, on the average, are to be found in subjective categories containing from four through seven words. Additional calculations indicate that from trial two on the average number of subjective categories used by each S is approximately fifteen for these 72 words.

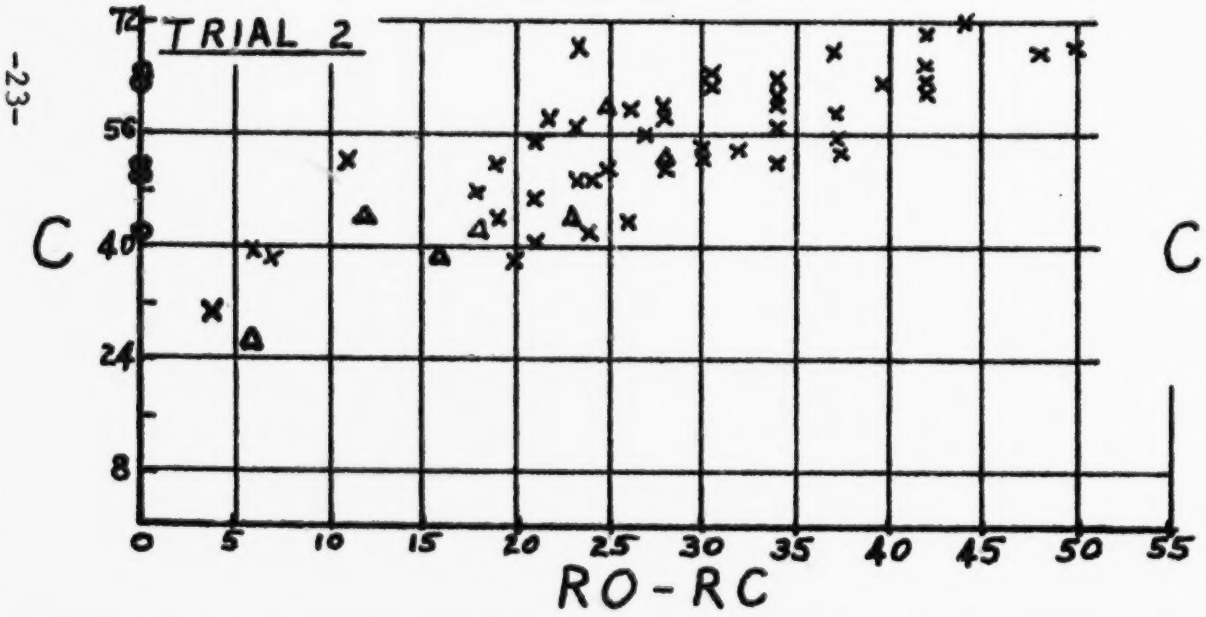
Fig. 9 depicts the relation between the degree to which individual

- - - - - - - - - - - - - - -
Insert Figure 9 about here
- - - - - - - - - - - - - - -

Ss utilized their own subjective organization (on the abscissa) and the number they get correct on a test (on the ordinate). The independent variable is the same index used previously, namely, repetitions observed minus repetitions expected by chance (RO - RC). However, the clusters are here defined in terms of the Ss' own study sheets. Several things are obvious from the scatter plots. First, ignoring for the moment those points falling on the chance (or zero) value of the independent variable, a relatively strong correlation is depicted between the two measures. In trial two, for example, this correlation approaches 0.8. In trials three and four, an obvious ceiling effect is present and the correlations consequently decrease. The points at the zero or chance clustering line are primarily for those Ss who showed no scoreable organization on their study sheets, that is, they were order of presentation arrangements. It is fairly obvious that as learning takes place (test 1 to 2 to 3, etc.) the numbers of these points decrease quite markedly. In general, proceeding from trial one to two to three to four the points move from left to right indicating a growth in subjective clustering as learning takes place. Thus, we have number correct associated with degree of clustering across Ss, and number correct associated with degree of clustering within Ss across trials. In summary: the more a S utilizes his own subjective clustering in ordering the words on his test, the better he is likely to do in terms of number correct.







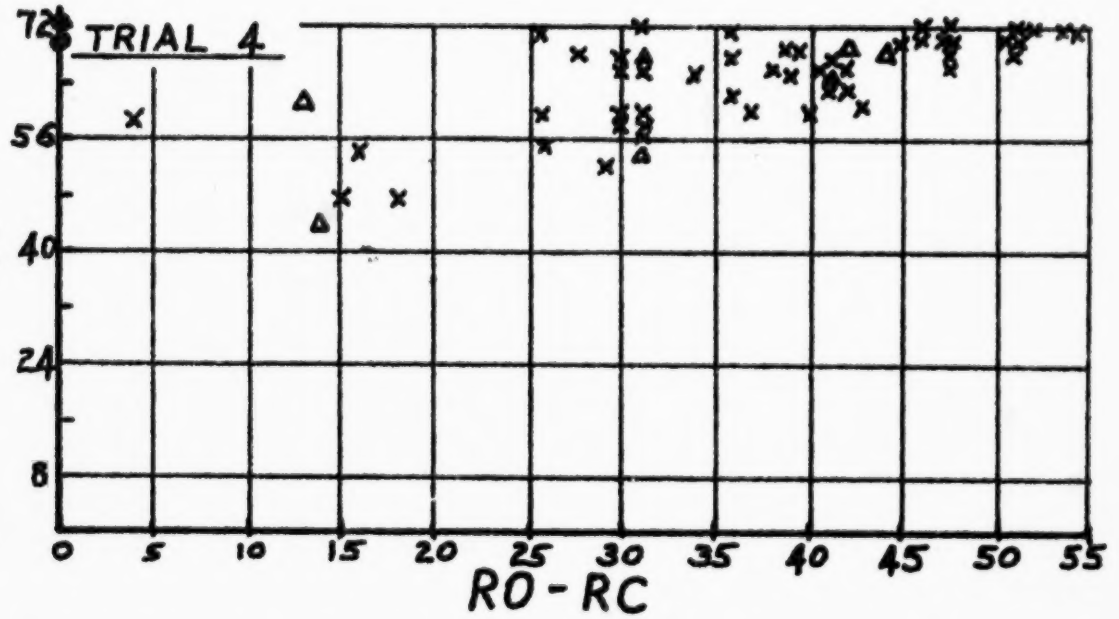


Fig. 9. Number of words correct (C) as a function of degree of subjective organization on the test (RO - RC), with subjective organization defined in terms of the individual study sheet. X's represent those Ss who organized according to meaning on their study sheets, triangles represent the alphabetic organizers, and circles represent those Ss who wrote the words on their study sheet in the order of presentation.

Another way of measuring subjective organization is in terms of the consistency of word order from one test to the next for each S. Tulving (1962) and Carterette and Coleman (1963) utilized what is referred to as the SO measure for this prupose. The present analysis utilizes the Kendall Tau coefficient, a measure very similar to the rank order correlation coefficient.

Fig. 10 depicts frequency distributions or hystograms for the Kendall Tau coefficients calculated between adjacent trials. The distributions are indicated separately for those Ss who organized in terms of the meanings of the words (the "a" hystograms). The "b" hystograms

- - - - - - - - - - - - - - - -
Insert Figure 10 about here
- - - - - - - - - - - - - - - -

include those Ss who wrote words on their study sheets in the order in which they were presented for at least one trial, and the "c" hystograms are for those Ss who utilized alphabetic organization for at least one trial. Thus, there are three hystograms for each adjacent pairs of trials. In general there is no marked movement toward higher Tau coefficients with learning, except for the "b" hystograms in going from trial one-two to trial two-three. Even for the last pair of trials (Fig. 10) the Tau coefficients are generally moderately positive but not very impressive. Some alphabetizers exhibit near perfect order co-relations, that is, Taus near 1.0, but there are also alphabetizers with lower Taus. Some alpha organizers simply cluster in terms of common first letters of words making no attempt to order these clusters alphabetically on their tests. These Ss account for the low Tau coefficients among the alphabetizers. In summary: a tendency for S to write words in the same order from trial to trial shows no growth as learning takes places, with the possible exception of a very small sub-group of alphabetizers who fully utilized alphabetical organization. It is important to note, of course, that the Kendall Tau coefficient is not the same index of consistency of sequence that Tulving (1962) and Carterette and Coleman (1963) used. Consequently, the results are not directly comparable.

Is there an optimum number of words for a category in order to minimize errors? Fig. 11 examines the relation between proportion of words in error and the size of subjective category in which those words were found. There is a subfigure for each of the five trials. Category

- - - - - - - - - - - - - - - -
Insert Figure 11 about here
- - - - - - - - - - - - - - - -

size on the study sheet is plotted on the abscissa. On the ordinate there is a ratio, the numerator of which is the proportion of words missed (P_i) on the test for the subjective category of size i, and the denominator is the size of the category i. P_i for the category of size two (P_2), e.g., is found by tabulating, for each S on a given trial, the number of words in two-word subjective categories which were missed

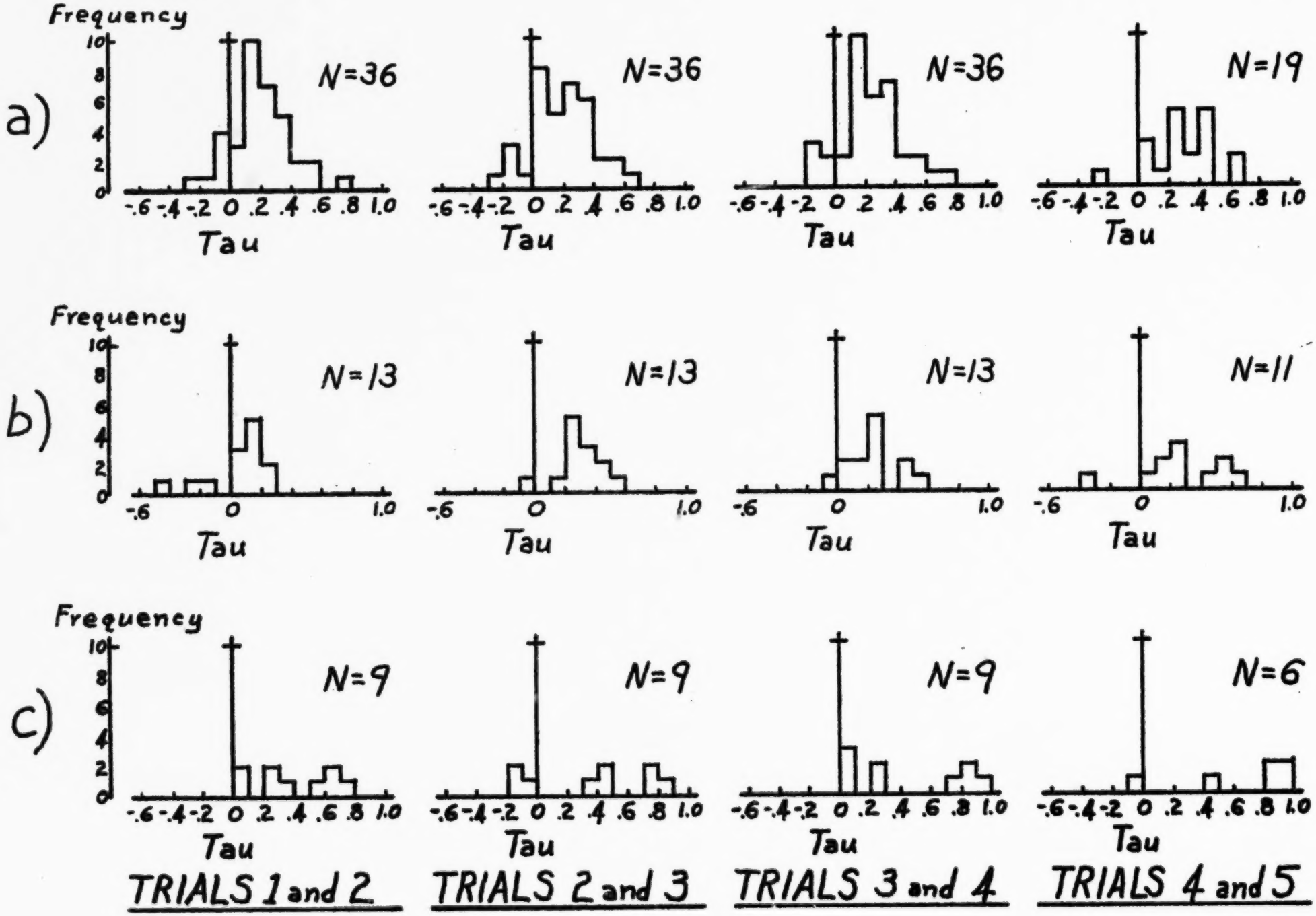


Fig. 10. Frequency distributions of Tau coefficients for test word-order on adjacent trials for a) those Ss who organized according to meaning, b) those Ss who wrote words in the order of presentation on at least one study sheet, and c) those Ss who alphabetized on at least one study sheet.

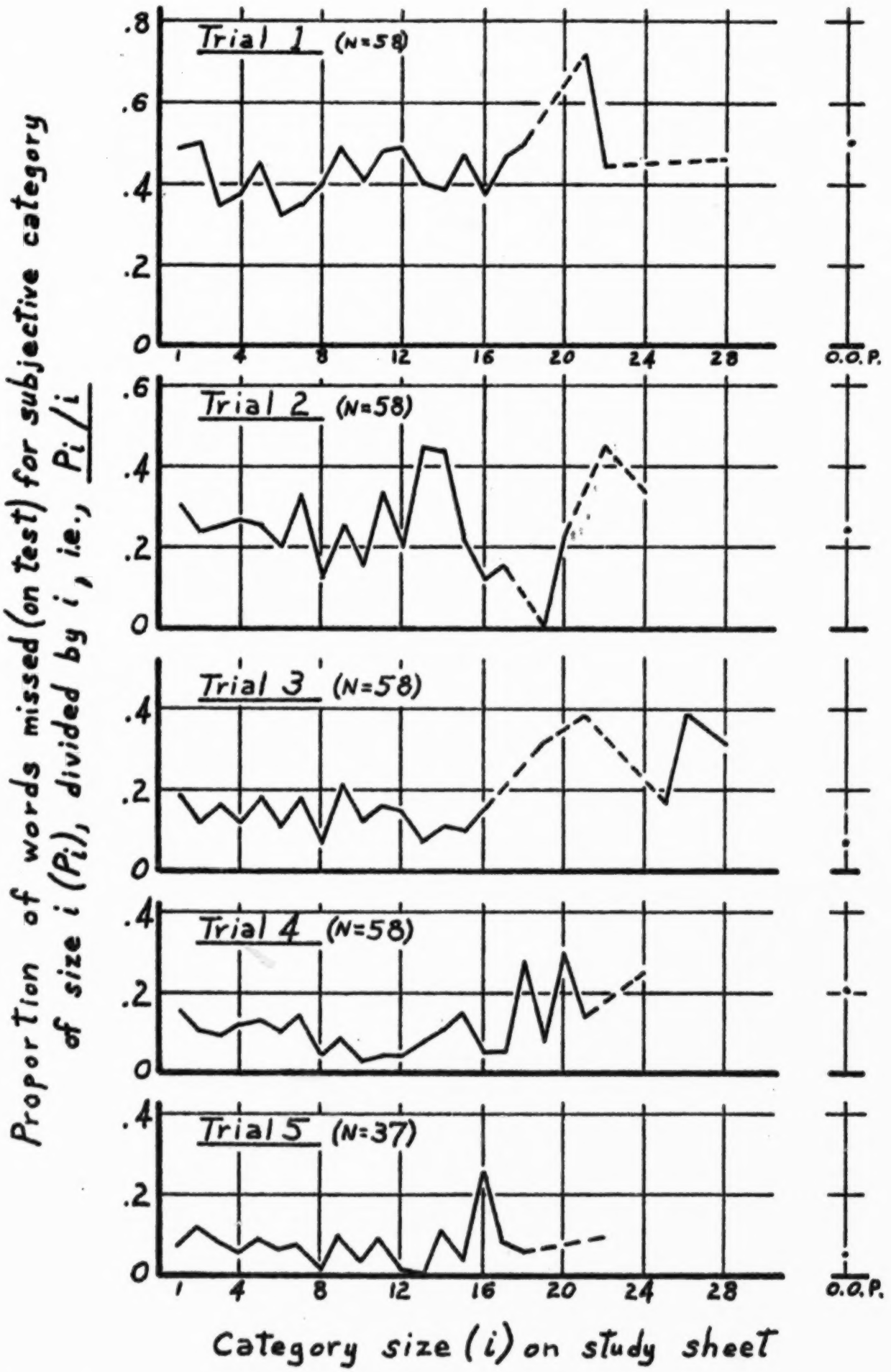


Fig. 11. Proportion of words missed for a subjective category of size i (P_i) is approximately equal to a constant, reflecting the overall average error rate (P), times i, i.e., $P_i \simeq (P) \cdot (i)$, or $(P_i)/(i) \simeq P$. There appear to be no "optimum" subjective category sizes. (O.O.P. stands for Order Of Presentation).

on the corresponding test. This number wrong is then summed across Ss yielding a total number wrong for category size two (W_2). Also tabulated is the number of different categories of two words used by each S. This number-of-times-that-categories-of-size-two-are-used is then summed across Ss to yield a count of the total number of times a category of size two was used (U_2). P_2 equals the ratio of (W_2/U_2). Dividing P_2 by 2 (in general, P_i by i) is, in effect, multiplying U_2 by 2 (or, in general, U_i times i). The resultant ratio, the one on the ordinate of Fig. 11, is one of number of words wrong divided by number of words involved. The general picture in Fig. 11, it is proposed, is one of a horizontal line with a fair amount of noise. Number wrong divided by number of words involved is a constant (P) across the different category sizes, i, where the constant (P) is simply the overall average error rate. The implication is, simply, that there appears to be no "optimum" subjective category size. If there were, one should find a dip in the curves, consistent across trials, in the region of the optimum value of i. While these findings are not to be directly compared with those of Dallet (1964), for example, because of the distinction between subjective category size and experimenter-defined category size, they are indirectly in support of the conclusions reached by Cohen (1963 and 1966) concerning a constant proportion of category recall across category sizes.

New Measures of Subjective Clustering

Additional measures have been developed for characterizing several additional aspects of Ss' subjective organization. The first of these measures (CON) provides an index of the consistency of an S's subjective organization from one trial to the next. The second measure (STR) provides an index of stereotypy of organization, or the degree to which words are grouped in the same way--across Ss. The third measure (CCP) is similar to the stereotype measure, but is designed specifically for examining selected pairs of words of particular interest to E. It reflects the proportion of Ss who put both words of the pair into the same subjective category, with the proportion "adjusted," in effect, for the sizes of the categories involved. The mnemonic stands for "common categorizing of pairs."

All three measures utilize the same basic concept which is, in essence, a "common elements" definition of the square of the correlation coefficient. Fig. 12 illustrates the basic form of the measure. Consider,

- - - - - - - - - - - - - - -
Insert Figure 12 about here
- - - - - - - - - - - - - - -

e.g., Roman numerals I and II as trials I and II. A S grouped or clustered words B, C, & D on trial I, leaving words A and E as isolates. On trial II his organization of the words shifted as indicated, i.e., A and B were in one group, and C, D, and E in another. Thus, each word appears in

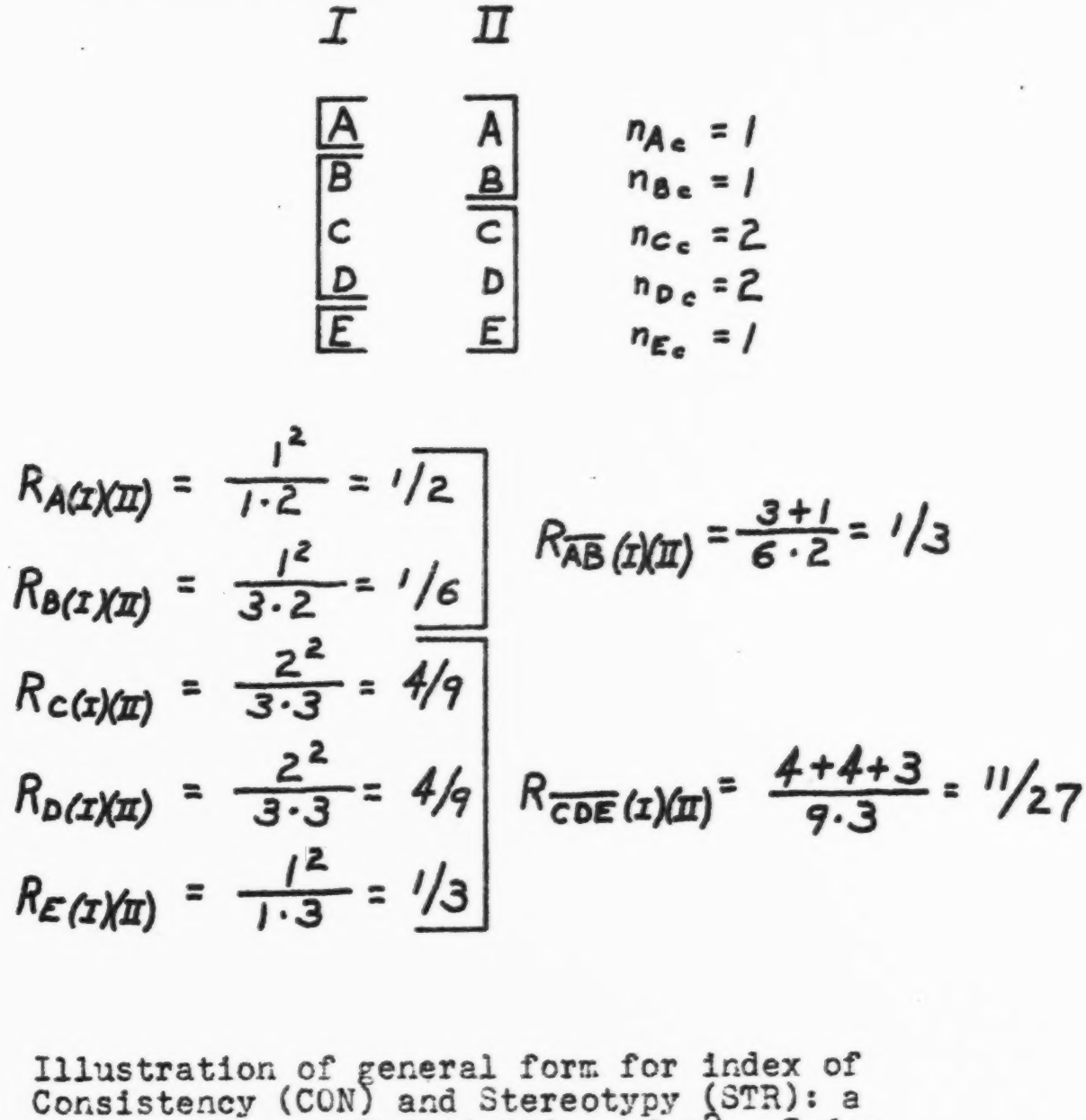


Fig. 12. Illustration of general form for index of Consistency (CON) and Stereotypy (STR): a "common elements" definition of r^2. Index of Common Categorizing (CCP) for a pair of words is based on the proportion of Ss who put both words in the same category, with the proportion "adjusted" for size of categories.

two groupings or clusters, one for trial I and the other for trial II. We may now look at each of the words, one at a time, and ask about the overlap of the groups in which they are to be found. The two groupings containing word C, e.g., have two words or elements in common, i.e., words C and D. Thus, the number of common elements for the two groupings is two. This number goes into the numerator and is squared. The denominator simply contains the product of the number of elements in each of the two groupings which contain C. The result is $R_{C(I)(II)} = 4/9$. This index is calculated for each of the words which appear on both lists. These indexes are labelled $R_{A(I)(II)}$ through $R_{E(I)(II)}$ in the illustration in Fig. 12.

For the consistency measure (CON) the index is simply averaged across all of the words of interest. This may be for particular pairs or groups of words (e.g., $R_{\overline{AB}(I)(II)}$ and $R_{\overline{CDE}(I)(II)}$ in the illustration of Fig. 12), or for all of the words on the list. To date only the latter, or overall, measure has been examined in detail.

The sterotypy index (STR) is also very simple conceptually, though it takes a computer to accomplish the very large number of calculations. Returning to Fig. 12 consider Roman numerals I and II as representing a pair of Ss on a given trial, instead of two trials for a given S. This is the basis of the index. The computer program then calculates one such index for each of the words, for every possible pair of Ss (the program will handle up to 20 Ss at a time), and then averages across all of the pairs of Ss. These averages may then be examined for individual words, or further averaged across pairs or sets of words.

The index of the common categorizing of pairs of words (CCP) is based on a count of the number of words in the categories within which a particular pair of words (e.g., A and B) may be found. For a given trial (study sheet) this count is made for each S and summed across the set of N Ss. The square of the sum goes into the numerator of the index. The denominator contains the product of two sums, one for the number of words in the categories within which word A is found and the other the corresponding sum for word B.

Details and illustrative computations are contained in Appendix A.

In the initial evaluation of the three new measures it was necessary to find twenty Ss with appropriate data. The study-sheet paradigm does not guarantee useable data for every S. Only those Ss could be used who exhibited some form of "meaning" organization, that was scoreable, on every trial. It was necessary to delete from consideration two Ss for not following instructions, two Ss who adopted an alphabetizing strategy, two Ss who wrote the words on their study sheet in the order in which they were presented for one or more trials, and three Ss for whom the scoring rules failed on one or more trials (indicating more than 1/2 of their words were isolated single words). Thus these initial analyses are based on the first 20 (out of 29) "good" Ss. Similar analyses with additional Ss are in progress.

Results of Experiment Two with Respect to the New Measures of Subjective Clustering

Fig. 13 is a series of four scatterplots for the 20 Ss, relating the index of consistency (CON) and the number of words correct. The plot

- - - - - - - - - - - - - - -
Insert Figure 13 about here
- - - - - - - - - - - - - - -

in the upper left, i.e., trial one to two study sheet consistency on the abscissa and number correct on the trial two test on the ordinate, depicts the only significant r. $r = 0.54$ ($.01 < p < .05$).[2] It appears as if number correct is correlated with consistency of organization only for performance very early in learning. There is a suggestion in the plots that a ceiling effect may be washing out the correlation in the later trials. Fig. 14 is included to argue against that interpretation. The

- - - - - - - - - - - - - - -
Insert Figure 14 about here
- - - - - - - - - - - - - - -

dependent variable is the same, i.e., the number correct. The independent variable is the index (RO - RC) which reflects the degree to which the individual Ss study sheet organization is reflected in his grouping of words on his test. The ceiling effect as trials progress does appear to be reducing the r's, but the r's are strong at least through trial three ($r = 0.74$, $p < .01$). The data depicted in Fig. 14, incidentally, very nicely replicate prior data on different Ss. For both Fig.'s 13 and 14 one may see a movement of the points to the right (and upward, of course) as trials progress. Consistency of study-sheet organization increases with trials, as does the degree to which the test organization reflects the study sheet organization. But, study sheet consistency appears to be related to number correct, across Ss only very early in learning.

The relationships between stereotypy (STR) and the E-defined variables of MR and categorization are depicted in the upper series of plots of Fig. 15. Stereotypy for the categorized pairs of words was consistently above that for non-categorized pairs. There is a general

- - - - - - - - - - - - - - -
Insert Figure 15 about here
- - - - - - - - - - - - - - -

downward trend in stereotypy in going from Hi MR to Lo MR, though there is a slight upturn at the two lowest MR levels. Stereotypy seems to first increase and then decrease across trials.

The relationships between the index of common clustering (CCP) for the E-defined pairs and the E-defined variables of MR and categorization are depicted in the lower series of plots of Fig. 15. The relationships

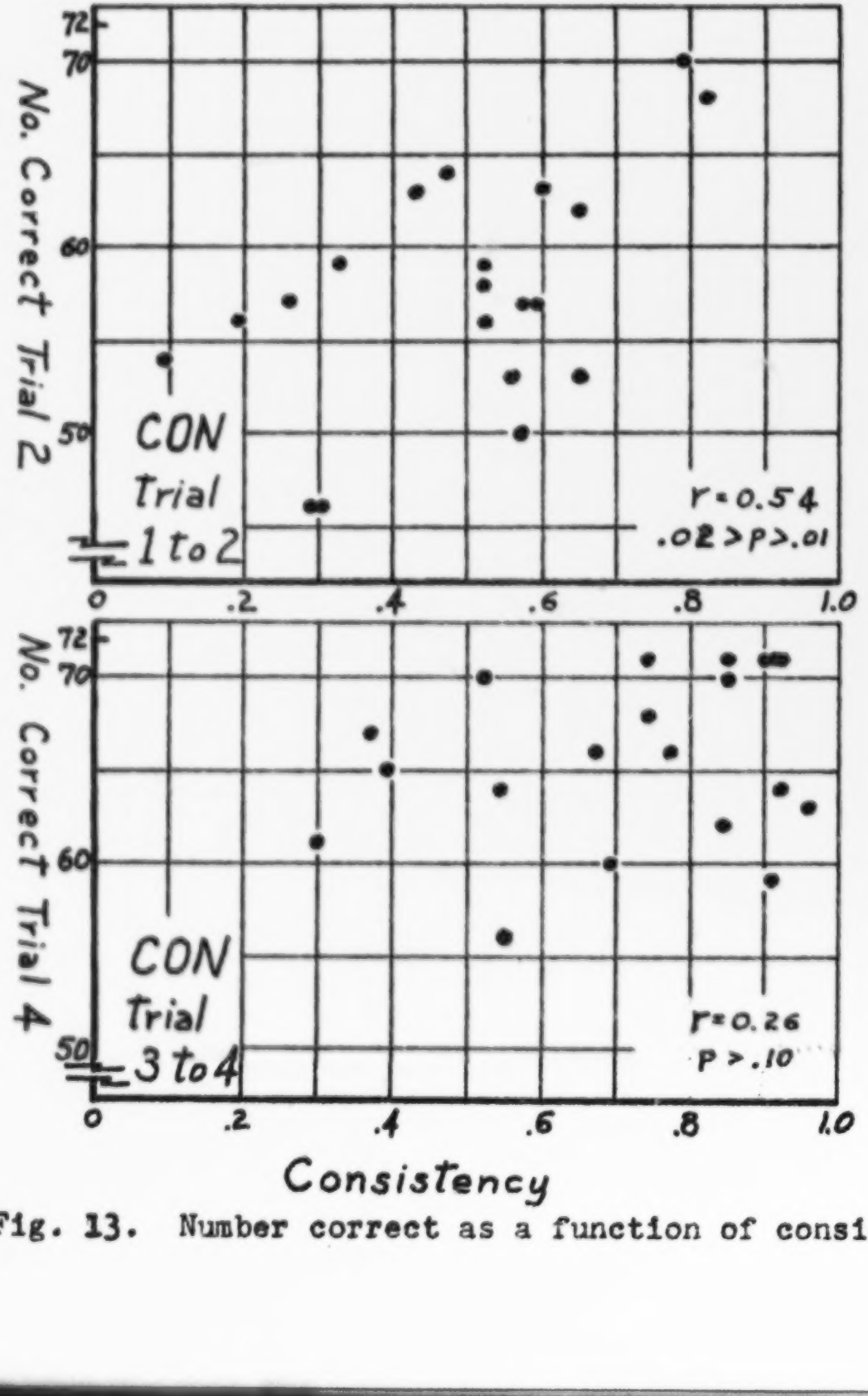


Fig. 13. Number correct as a function of consistency of organization (CON).

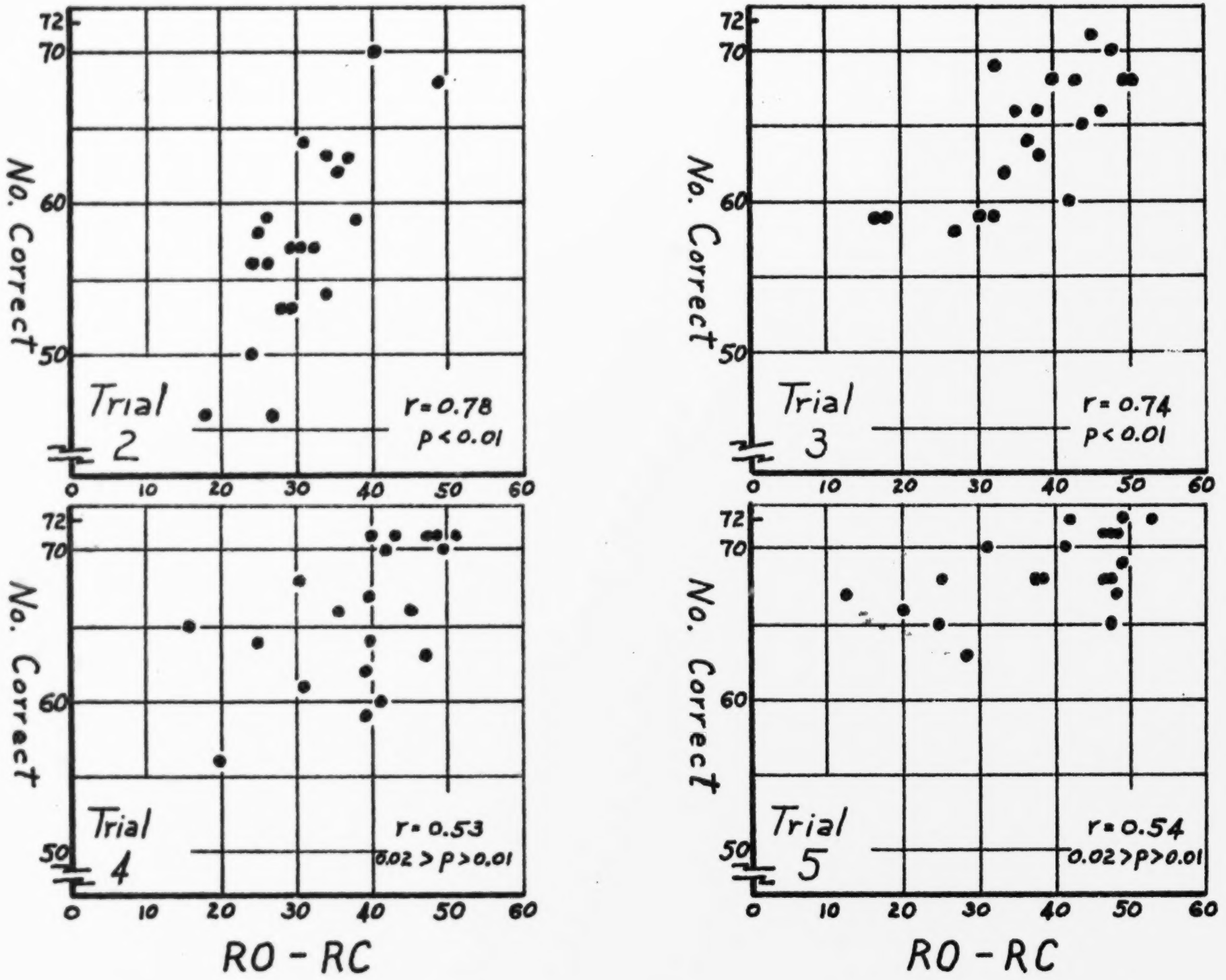


Fig. 14. Number correct as a function of degree of subjective organization (RO - RC).

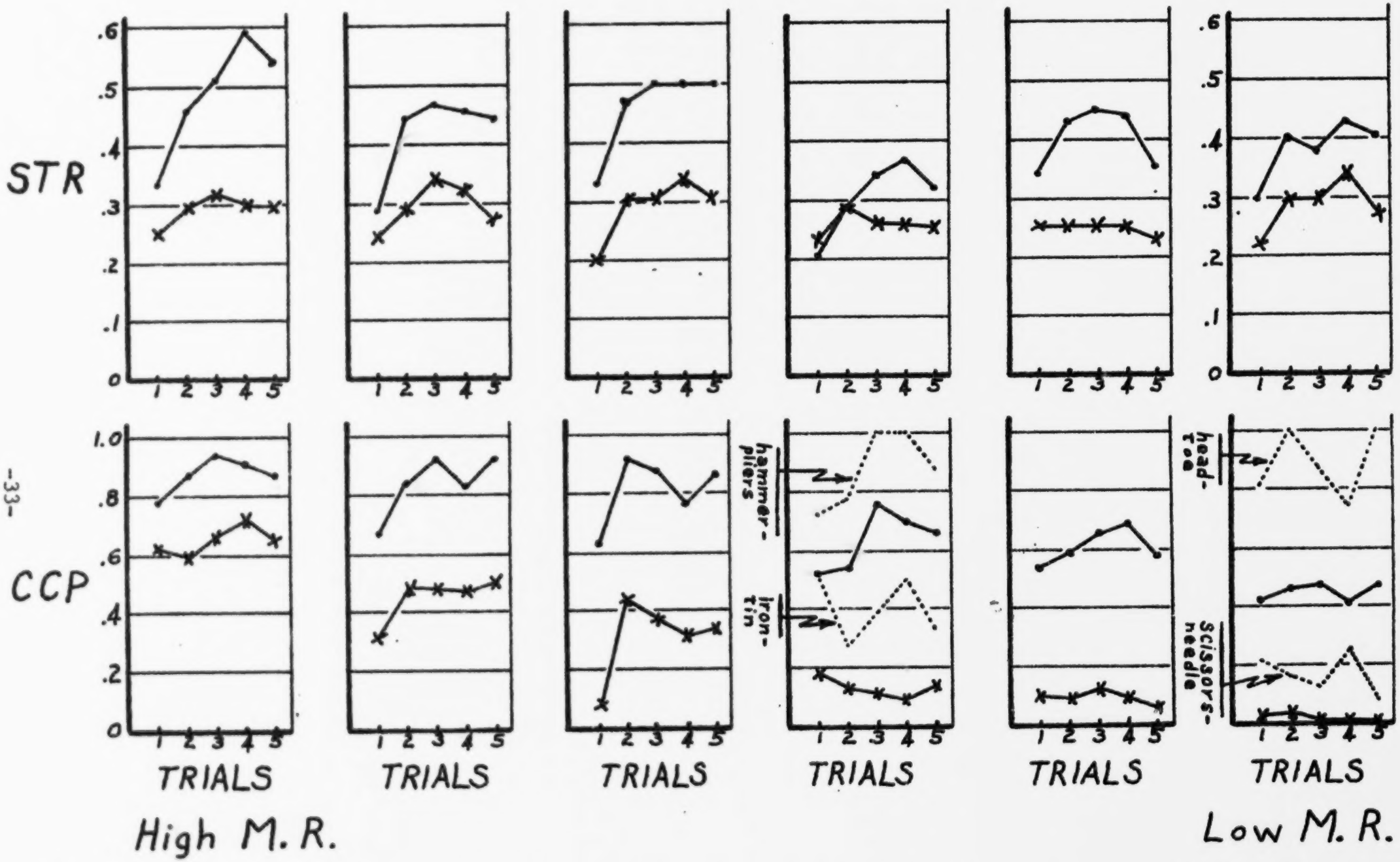


Fig. 15. Stereotypy of organization (STR) and common categorizing of pairs (CCP), for Categorized (•——•) and Non-Categorized (×——×) pairs of words at different Mutual Relatedness (M.R.) levels, as a function of trials. The dependent variables are based on study sheet performance.

are more orderly than for the STR measure, and nicely parallel the relationships depicted in Fig. 7 for the (RO - RC) measure. The difference between categorized and non-categorized pairs is smallest for high MR values, and CCP decreases with decreases in MR. These results also parallel those of the Marshall (1967) study referred to earlier. Examination of individual pairs of words, however, indicates that there are still some "big chunks of variance" to be accounted for. Some of these are indicated as dotted lines in the bottom portion of Fig. 15. Considering the lowest MR plot, the dotted line that reaches 1.0 represents the word pair "head-toe." The very low dotted line is for "scissors-needle." Both of these word-pairs are members of the categorized-lowest MR group. There is more difference between these two pairs than there is across the entire MR range for categorized pairs (though, of course, these extreme pairs were picked to exaggerate the point). It turns out that almost every S adopted a "parts-of-the-body" subjective category. Perhaps this was triggered by the higher MR pairs "arm-leg" and "foot-knee" which were also in the list. At any rate, most Ss put "head" and "toe" in the same subjective category despite the low MR value relating them. In contrast, "scissors" frequently went into a subjective category with "hammer" and "pliers" and/or "dagger," while "needle" often went with "silk," "dress," "glove," and "cloak." Very few Ss put "scissors" and "needles" in the same subjective category. These effects of the total context of the list cannot be ignored if these "big chunks of variance" are to be accounted for.

Discussion for Experiment Two

A completely objective procedure for defining the subjective clusters of Ss during learning has been worked out. This procedure depends completely on the geometry of the filled and empty cells on a study sheet. Judgments concerning "what Ss intended" are not necessary. Though the objective procedure has been worked out it is time consuming and expensive. It is suggested that much of this time and expense can be saved with very minor modifications to the study sheet paradigm. Instead of giving the Ss complete freedom as to how to organize on their study sheets, the Ss are to be instructed to place those words which they wish to go together into the same column. With an illustration, but with care to avoid suggesting any particular organization, it should be possible to get Ss to continue to "organize the words on your study sheet so as to best help you to memorize." Assuming that Ss follow instructions it will then no longer be necessary to go through the expensive and time consuming objective procedure for determining whether a particular S organized by columns or by rows.

The two E-defined variables which were built into the stimuli, i.e., the MR strength of word-pairs and the distinction between categorized and non categorized word-pairs, were quite potent with respect to subjective clustering. This was true in terms of the (RO - RC) measure, the STR measure, and the CCP measure. Word pairs with high MR strength are salient and persuasive. Most Ss utilize these pairs in their subjective organizations. Over all MR levels however, the E-defined categories show

a relative decrease in frequency of usage as learning progresses when compared with the frequency of usage of the subjectively defined categories. Even for the highest MR levels clustering behavior in terms of the E-defined categories never comes close to complete or perfect clustering.

In terms of the E-defined variables built into the stimuli there is strong agreement across the different measures of this study, namely; (RO - RC), CCP, and to a considerable extent STR. These general results also agree to a very marked extent with the comparable results of the Marshall (1967) study despite wide differences in procedure, measurement, design, and S population.

Ss use an average of approximately 15 clusters for these 72 words. Clusters get larger with trials. Most of the 72 words are contained in the clusters of from three to eight words each. There is no optimum size of a subjective cluster in terms of number of errors made. It is also proposed that there is no optimum number of categories either. This is in contrast to the report by Mandler (1967). A preliminary examination of the data of the present study, via scatter plot, indicates no relation between the number of categories used and the number of items correct. A careful evaluation of these data is yet to be made. "Goodness" of a subjective chunk does not depend on the number of items in it. And, the total list may be thought of as one big chunk with the categories within it as its elements. In the conditions of the study-sheet paradigm Ss manipulate the word elements until they achieve subjective chunks which are approximately equally "good" subjectively. "Good" of course means easily learned and (yet to be evaluated) remembered.

Ss whose tests reflect their study sheet organizations are the ones who learn fastest. The degree of correspondence between study and test organization increases with trials for almost all Ss.

Consistency of word order from one test to the next, as measured by Kendall's Tau, is only moderately positive and shows no consistent growth over trials. Consistency of organization from one study sheet to the next does show growth over trials, but for any one trial it is related (across Ss) to number correct only very early in learning.

The three new measures developed permit measurement of aspects of subjective organization which have not been measured before. The stereotypy and common categorization measures have been shown to have a kind of concurrent validity in terms of the E-defined built in variables. Validity for the consistency measure is not as clearly established as yet.

The common categorization measure has pointed out the need to take

the total list context effects into account in accounting for subjective organization behavior and, by inference, learning. These effects are a big source of variance.

Experiment Three

The stimuli for this experiment were designed to be very salient and persuasive with respect to organization. Exhaustive categories of words were used, e.g., north, sough, east, and west, or mother, father, sister, and brother. Most of the words were taken from Cohen's (1963) report. The complete set of stimuli is shown in Table 3. Original plans were to continue to utilize the study-sheet paradigm in order to check

- - - - - - - - - - - - - -
Insert Table 3 about here
- - - - - - - - - - - - - -

on the persuasiveness of the E-defined categories. However, a pilot study indicated that the categories were indeed very persuasive, and that most Ss learned most of the 70 words in just over one trial. For these reasons the study-sheet paradigm was abandoned for this experiment. It was assumed that almost all Ss would utilize almost all of the E-defined categories.

Data reported by Cohen (1963 and 1966) indicate that if an exhaustive category is remembered at all (i.e., at least one member of it) then most of that category will tend to appear on a free recall test. From this, one may reason that the total of individual items of an exhaustive category "come out" because of the pre-experimental history of S, and only one item, or the name of the category, need be learned during the experiment. If one item is eliminated from each of the exhaustive categories, however, then S must also learn during the experiment which item of each category has been left out. Hence, if other things could be kept equal, one would predict "the shorter" list, i.e., the list with one item missing from each category, should be the more difficult list to learn. But it is not possible to keep "other things equal:" a) if one eliminates one item from each category then the stimulus list has been shortened by as many words as there are categories in the list. Shorter lists are easier to learn (in terms of proportion correct) than longer ones, but the categorization effect predicts that the longer list will be easier. The eventual direction of the difference will then be a function of which of the two factors is more potent. b) If an attempt is made to keep the lists the same length then categories must be added to the incomplete-category-list, and then the category effect would be confounded with number of categories. Under these conditions both effects would be expected to operate in the same direction. Hence, alternative (a) was chosen. It was hypothesized that the category effect would be more potent than the difference between a 70-item list and a 50-item list thus predicting the 50-item "incomplete-category" list more difficult to learn. This is a dramatic, counter-intuitive prediction: a given list will be easier to learn than the same list with 29% of the words removed. Since there really was very little evidence to support the prediction

TABLE III. Stimulus words for Experiment 3.

Words with ** were eliminated for both 50- and 60-word lists, and words with * were eliminated for just the 50-word list.

*ALTO	EAST	*FALL	FOOT	COCAIN
BASS	NORTH	SPRING	*INCH	HEROIN
SOPRANO	SOUTH	SUMMER	MILE	*MORPHINE
TENOR	*WEST	WINTER	YARD	OPIUM
ACE	BROTHER	DIME	FRESHMAN	ADDITION
**JACK	**FATHER	NICKEL	**JUNIOR	DIVISION
KING	MOTHER	**PENNY	SENIOR	**MULTIPLICATION
QUEEN	SISTER	QUARTER	SOPHOMORE	SUBTRACTION
*BREAKFAST	*CAUCASIAN	FORK	*FEMININE	BEER
DINNER	NEGRO	KNIFE	MASCULINE	WHISKEY
LUNCH	ORIENTAL	*SPOON	NEUTER	*WINE
**AMPERE	ANIMAL	CIGAR	MARINE	BLONDE
OHM	**MINERAL	CIGARETTE	SAILOR	**BRUNETTE
VOLT	VEGETABLE	**PIPE	**SOLDIER	REDHEAD

concerning the relative potency of these two effects, a third condition was also included in the experimental design. For this condition a mixed list was presented to the Ss, providing a within-S comparison. Ten out of the 20 categories in the list were left complete and the other ten had one item removed from each of the categories.

Method for Experiment Three

A 70-word list was constructed which contained 20 exhaustive categories, ten categories with three words each and ten categories with four words each. (See Table 3). A group of 20 Ss learned this list. A second list was generated by removing one word from each of the 20 categories of the original list, thus leaving a 50-word list. A group of 22 Ss learned this list. The third list had one word removed from each of ten of the twenty categories, thus making for a 60-word list. A third group of Ss (n = 30) learned this list. For all three groups words were presented in random orders two to three seconds per word in a standard free recall paradigm for six trials. Five minutes was allowed for each free-recall test. The particular words eliminated for the 50- and 60-word lists are indicated in Table 3.

Results for Experiment Three

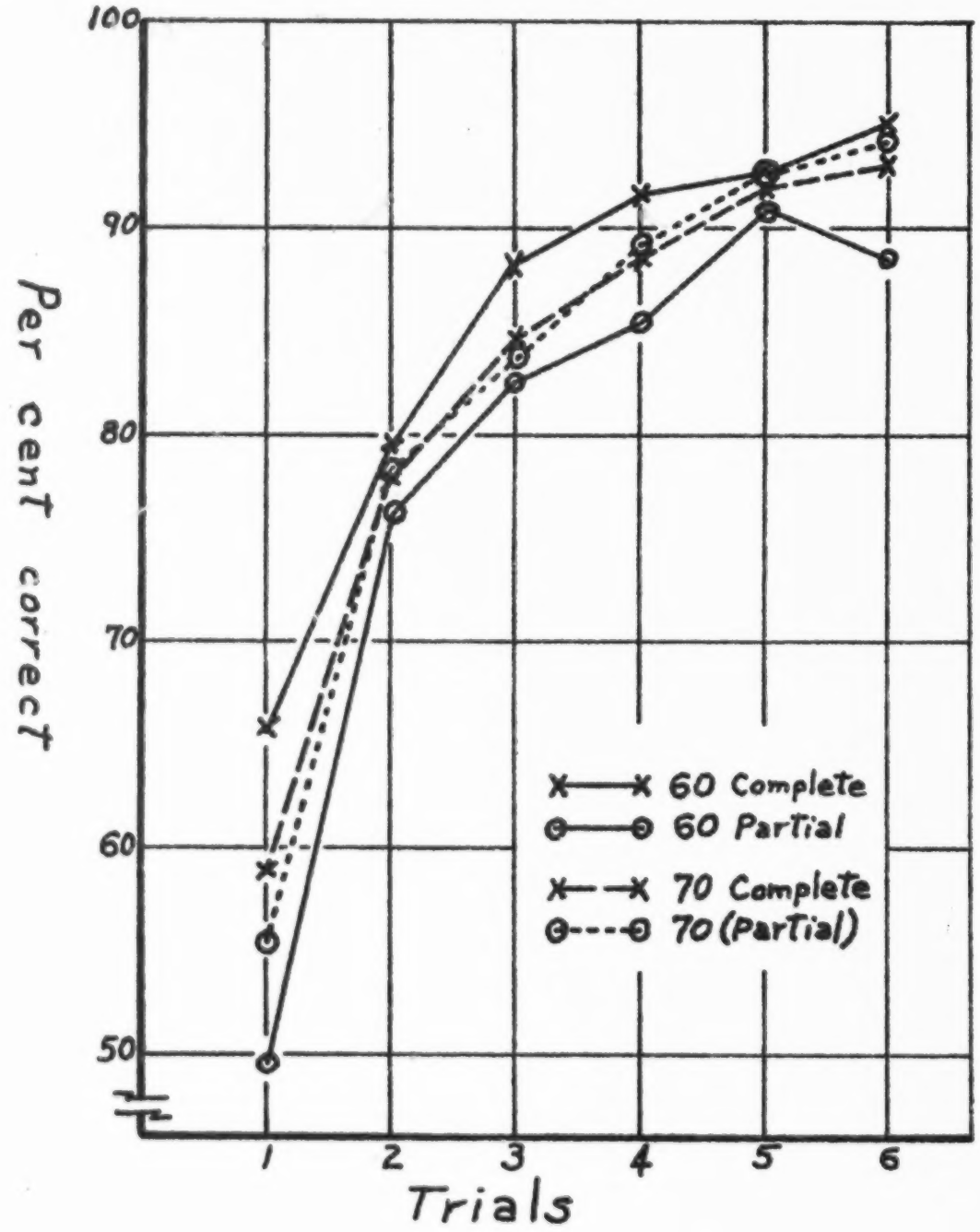
The percent of words correct for the complete and partial categories was the dependent measure. Results are depicted in Fig. 16. The definition

- - - - - - - - - - - - - - - -
Insert Figure 16 about here
- - - - - - - - - - - - - - - -

of "complete" and "partial" is obvious for the 60-word list. For the 70-word list all categories were complete, but the (partial) refers to those categories which were partial for the 60-word Ss. The "complete" were, of course, the same for both the 70- and 60-word lists. The (complete) and partial categories of the 50-word lists are defined in a similar way, i.e., with reference to what happened to those categories on the 60-word list.

The between-Ss evaluation of the major hypothesis of the experiment involves a comparison of the overall percent correct on the 50-word list with the equivalent measure on the 70-word list. Contrary to prediction performance on the 50-word list was higher than on the 70-word list, i.e., the two dashed curves, combined, in the right hand plot of Fig. 16 are higher than the two dashed curves, combined, in the left hand plot. If both categorization effect and length of list effect were operating, apparently the length of list effect was more potent.

The solid lines of Fig. 16, which are repeated in both left and right hand plots, depict the performance of the 60-word group. The percent correct for the complete categories (x's with solid lines) is significantly greater ($p < 0.01$) than for the partial categories (circles with solid lines).



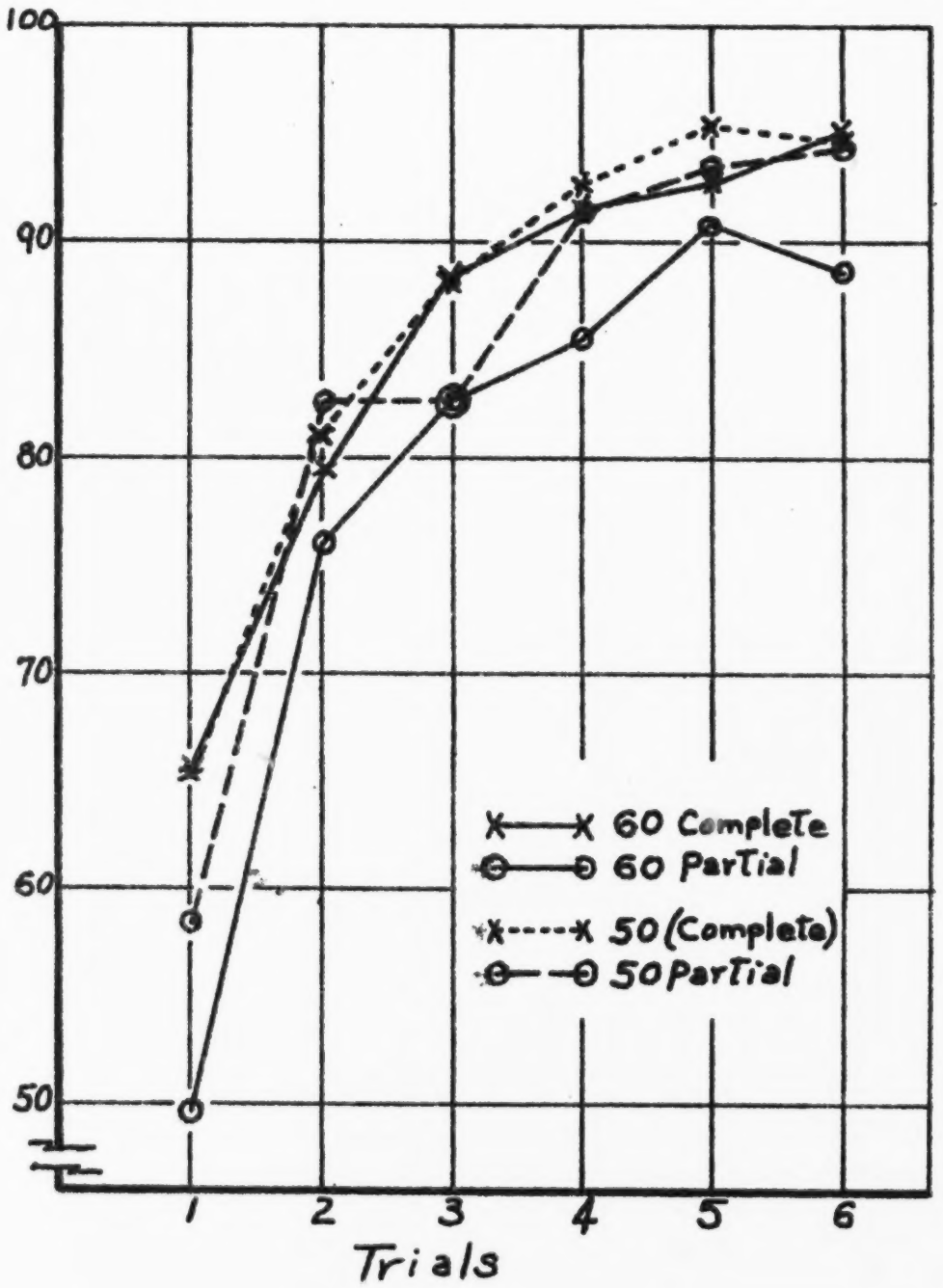


Fig. 16. Learning of lists of twenty exhaustive categories as a function of trials: one item missing from each category in the 50-item list; one item missing from ten of the categories in the 60-item list; and all items present in the 70-item list.

Thus the within Ss comparison does reveal a significant categorization effect, i.e., complete versus partial comparison. When these subsets of categories are compared for the 70-word and the 50-word Ss there is no significant difference ($p > 0.05$) between them. Thus the difference found for the 60-word group cannot be ascribed to the particular subsets of categories that were used as partial and complete. The interaction between trials and the difference between partial and complete is also significant ($p < 0.01$) for the 60-word group. An examination of the means, however, indicates no orderly progression for the differences between complete and partial across trials. The equivalent interactions for the 50-word and 70-word groups were not significant ($p > 0.05$)

The overall percent correct for the 50-word list (all categories incomplete) was slightly higher than that for the 60-word (mixed) list. However the overall percent correct comparison for the 60-word list with the 70-word list (all categories complete) indicates essentially identical performance for the two lists (groups). Comparisons between a mixed list condition (one half of the categories complete and one half incomplete) and the pure list conditions (all categories complete or all categories incomplete) confounds the mixed list versus pure list effect with the complete versus partial category effect, and both of these are confounded with number of words in the list. The interactions amongst these effects may be quite complex. However, the comparison of the 60-word and 70-word lists (left plot in Fig. 16) provides some interesting suggestions. Overall performance in terms of percent correct was essentially equivalent for the two lists. Yet, an examination of the complete versus partial category subgroups of the list shows: first, that there is no difference between these subgroups for the 70-word list (all categories complete); second, the identical categories (those which were complete) in the 60-word (mixed) list were easier to learn in the mixed list than in the 70-word list; while third, the incomplete categories of the 60-word (mixed) list were more difficult to learn than their complete versions appearing in the 70-word list. It is as if the Ss of the 60-word (mixed) list conditions concentrated first on the complete categories and only after obtaining some mastery of these did they switch their attention to the incomplete categories of the list. In the 70-word list there was no such differentiation amongst categories and both subsets of categories were learned with equal speed. The effects of mixed versus pure lists cannot be ignored in the evaluation of the basic hypothesis of this experiment.

Discussion for Experiment Three

The complete versus partial category effect has been shown for a mixed list, but the magnitude of the length of list effect (70 versus 50 words) has apparently swamped whatever category effect might have been operating in the between Ss comparison. A test of the major hypothesis with a between Ss comparison calls for lists which differ much less in total lengths. Going from a list of 70 words to one of 50 words is reducing the list length by 29 percent. Perhaps a reduction of only

five to 10 percent in list length would enable the category effect to be more potent and lead to positive results. In support of this (if one is willing to ignore the possible confounding effects of mixed versus pure lists) it may be noted that in going from 70 to 60 words, list length was reduced 14%, and there was no difference in overall percent correct for the two conditions. Several variations on the present experimental design may be suggested. One variation simply involves utilizing exhaustive categories with larger numbers of items, e.g., days of the week, months of the year, major cities in the state, etc. Removing just one member from each of these categories would reduce list length by a relatively small percentage. Another approach might involve an incomplete category list with one item per category deleted, thus deleting e.g., ten items from the list. The complete category list for comparison would have all but one item eliminated from enough of the categories so that the same total number of items, e.g., ten, were deleted. Thus, the comparison list would be of equal length in terms of total number of items, and would have the same number of categories represented. The difference would be that the first list would have, e.g., ten incomplete categories in it while the comparison list might only have one or two incomplete categories in it (these categories containing only one item each).

Conclusions and Implications

An experimental paradigm has been developed which enables one to study the subjective organizations which Ss impose on material that they are assigned to learn. Measures have been developed, based on the data provided by this paradigm, which permit evaluating the degree to which Ss utilize their own subjective organization during learning, the degree to which subjective organization is consistent from one trial to the next, and the degree to which subjective organizations are stereotyped across Ss. Three experiments, exploring various aspects of the organization of material during learning, permit the following general conclusions. Giving Ss the opportunity to overtly organize the material to be learned facilitates learning. Given the opportunity, those Ss who utilize it learn more than those who do not. The performance of those Ss who are given the opportunity to overtly organize, but who do not utilize that opportunity, is essentially indistinguishable from other Ss who were not given the opportunity to overtly organize the material.

E-defined categories of words, based on free-association normative data, vary considerably in the degree to which Ss perceive and utilize them in their subjective organization of the material. The concept dominance variable (Underwood and Richardson, 1956) is very weak in terms of persuading Ss to utilize the categories so defined in their subjective organization. The mutual relatedness variable (Cofer, 1965, and Marshall, 1967) is much more salient and persuasive. It covers the range from highly persuasive to not persuasive at all, but even at the highly persuasive end it is clear that these E-defined categories are still leaving large portions of variance unaccounted for in the subjective organization behavior of Ss. The context of the total list of items to be learned must

be taken into account before portions of this unaccounted-for variance will be understood.

When Ss subjectively organize material there appears to be no optimum size of subjective cluster for minimizing error. It is proposed that when Ss develop unrestricted subjective categories they are in essence manipulating the elements of the list until they achieve subjective chunks which are, for them, equally "good." "Good" in this context means easily learned. It is proposed that "good" may also be interpreted to mean "remembered" but retention data are not yet available.

Though Ss may strive to reduce all subjective categories to equal "goodness," there are characteristics of stimulus materials which make this relatively impossible. This difference in materials is ascribed to the very lengthy and extensive pre-experimental history which the S brings with him to the experiment. The exhaustive categories of experiment three are highly familiar to almost all Ss and almost all utilize them fully in their subjective organization of the material to be learned. However, an exhaustive category with one item missing is not as "good" as the corresponding complete exhaustive category. For the complete category S may utilize, without modification, his pre-experimental history. For the incomplete categories, however, S must add something to his pre-experimental learning in order to utilize it in the experiment.

Alexander and Huggins (1964) report on the use of an approach similar to the study-sheet paradigm for a perception experiment. Their results are also quite encouraging.

Summary

When presented with the task of learning meaningful verbal material (and other forms of material, as well) most Ss cluster or organize the items to be learned into subjectively meaningful groupings. An experimental paradigm (the "study sheet paradigm") has been developed which permits the objective measurement of this subjective clustering during learning. The unique subjective groupings for each individual S may be identified on each learning trial. The paradigm incorporates the use of specially prepared study sheets which each S prepares for himself on each learning trial. The information on these study sheets provides for the determination of the subjective clusters. Test performance may then be examined with respect to these subjectively defined clusters.

In addition to describing the study sheet paradigm the present report describes two experiments employing that paradigm, several newly developed measures for previously unmeasured aspects of the subjective organization of material, and a third experiment concerned with a particular implication of subjective organization behavior.

The material to be learned in the first experiment had groupings of words "built into it," with most of the groupings defined via the concept dominance data provided by Underwood and Richardson (1956). An experimental group of Ss learned the material via the study-sheet paradigm. One control group learned under identical conditions except that they did not have the opportunity to overtly organize the material on their study sheets. A second control group learned under "standard" free-recall conditions. The opportunity to overtly organize the material to be learned facilitated learning, as the experimental group achieved more words correct than either of the two control groups. Numbers of words correct for the two control groups were almost identical. Further, only those Ss of the experimental group who took advantage of the opportunity to overtly organize the material were the ones to show better performance. Performance for the Ss of the experimental group who showed no organization behavior on their study sheets was indistinguishable from the performance of the Ss in the two control groups. The category definitions based on concept dominance are not salient and/or persuasive. Ss tend to ignore them, and appear to find other criteria for categorizing or organizing the words. The concept dominance categories are reflected in test performance to a lesser and lesser degree as learning progresses. In contrast, the subjectively defined categories are reflected in test performance to a greater degree as learning progresses.

The material to be learned in the second experiment had pairs of words "built into it," with the pairs defined via the mutual relatedness data reported by Cofer (1965) and Marshall (1967). All Ss learned via the study-sheet paradigm. A new measure (CON) was developed for the consistency of organization from one trial to the next, independent of the order in which the words were written. A second new measure developed (STR) gave an index of the degree of stereotypy of organization for each word and/or the set of words, i.e., the degree to which the different Ss

organized the material in the same way. A third new measure (CCP) provides an index of the degree to which selected pairs of words are categorized into the same subjective categories, i.e., a measure of the common categorizing of the pairs. The mutual relatedness variable was considerably more potent than the concept dominance variable of experiment one. Word pairs with high mutual relatedness strength are salient and persuasive. Most Ss utilize these pairs in their subjective organizations. Over all, however, the "built-in" pairs show a relative decrease in frequency of usage as learning progresses when compared with the frequency of usage of the subjectively defined categories. Even for the highest mutual relatedness levels clustering never comes close to complete or perfect. In terms of the mutual relatedness variable there is strong and striking agreement between the results of the present study and those of Marshall (1967), despite wide differences in procedure, measurement, experimental design, and S population. There is no optimum size of a subjective cluster, and Ss use an average of approximately 15 clusters for these 72 words. Ss whose tests reflect their study-sheet organization are the ones who learn fastest. The degree of correspondence between study and test organization increases with trials for almost all Ss. Consistency of word order from one test to the next shows no consistent growth over trials, but consistency of organization (as measured by CON) from one study sheet to the next does show growth over trials. However, CON is related (across Ss) to number correct only very early in learning. The STR and CCP measures very nicely reflect the variables "built into" the stimuli, and thus exhibit a kind of concurrent validity. The CCP measure clearly points out the large source of variance due to total list context effects, and the need to understand these effects in order to account for subjective organization behavior and, by inference, learning.

The third experiment utilized lists of exhaustive categories (Cohen, 1963). For one group of Ss all categories were complete, for a second group one word was missing from each category, and for a third group one word was missing from one half of the categories. Exhaustive categories are highly salient and persuasive. The pre-experimental history of the Ss permits them to re-generate most of the words of the categories given only that they remember one of the words (or the category name). However, if one randomly selected word is omitted from a category, then S must also remember which word to leave out. It was hypothesized that the incomplete categories would be more difficult to learn than the complete ones. The between Ss comparison (groups one and two) was confounded with length of list, and apparently list length was the more potent variable. The results were counter to the hypothesis. However, the within Ss comparison (group three) clearly supported the hypothesis, and the data suggested that it should be possible to design a between-Ss experiment which would also support the hypothesis.

References

Alexander, C. and Huggins, A. W. F. On changing the way people see. Percept. mot. Skills, 1964, 19, 235-253.

Bousfield, W. A. and Puff, C. R. Clustering as a function of response dominance. J. exp. Psychol., 1964, 67, 76-79.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. A study of thinking. New York: Wiley, 1957.

Carterette, E. C. and Coleman, E. A. Some comments on Tulving's subjective organization paper at Psychonomic Soc. Meeting, Bryn Mawr, Pa., 1963.

Cofer, C. N. On some factors in the organizational characteristics of free recall. Amer. Psychologist, 1965, 20, 261-272.

Cohen, B. H. Recall of categorized word lists. J. exp. Psychol., 1963, 66, 227-234.

Cohen, R.H. Some-or-none characteristics of coding behavior. J. verb. Learn. verb. Behav., 1966, 5, 182-187.

Dallett, K. M. Number of categories and category information in free recall. J. exp. Psychol., 1964, 68, 1-12.

Mandler, G. Organization and memory. A chapt. in Spence, K. W. & Spence, J. T. (eds.) The psychology of learning and motivation: advances in research and theory. Vol. 1. New York: Academic Press, 1967.

Marshall, G. R. Stimulus characteristics contributing to organization in free recall. J. verb. Learn. verb. Behav., 1967, 6, 364-374, (espec. Exp. II).

Tulving, E. Subjective organization in free recall of "unrelated" words. Psychol. Rev., 1962, 69, 344-354.

Underwood, B. J. and Richardson, J. Some verbal materials for the study of concept formation. Psychol. Bull., 1956, 53, 84-95.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Appendix A

I. Index of stability of organization for word k from trial t-1 to trial t ($R_{k(t-1)(t)}$) for any given subject, i. With KK words in the list which is to be subjectively organized (words A, B, C, . . . k, . . . KK) build a one by KK matrix for each word, k, for trial t-1, and another for trial t. Considering all KK words this means, in essence, a square matrix KK by KK big for each trial. For each cell entry for word A (row A), for example, let the entry (c_{Ak}) equal one if word A <u>was</u> in the same subjective category as word k, let c_{Ak} equal zero if word A was <u>not</u> in the same subjective category, and let C_{AA} equal one. This is done separately for each trial for a given subject. Define $R_{k(t-1)(t)}$ as:

(1) $R_{k(t-1)(t)} = n_{kc}^2 / (n_{k(t-1)}) \cdot (n_{k(t)})$, where, for word A, e.g.

$n_{Ac} = \sum_{k=A}^{KK} (c_{Ak(t-1)}) \cdot (c_{Ak(t)})$, which is a count of the number of words categorized with word A on trial t-1 that were also categorized with word A on trial t (the minimum value is 1, for word A alone). This count is squared for the numerator

$n_{A(t-1)} = \sum_{k=A}^{KK} c_{Ak(t-1)}$, which is the number of words categorized with word A on trial t-1.

$n_{A(t)} = \sum_{k=A}^{KK} c_{Ak(t)}$, which is the number of words categorized with word A on trial t.

$R_{k(t-1)(t)}$ is interpretable as the square of a correlation coefficient defined in terms of "common elements." In other words, it is the product of two proportions, $n_{Ac} / n_{A(t-1)}$ times $n_{Ac} / n_{A(t)}$. If the number of common elements equals the total number of elements for either trial t-1 or trial t, then the index of stability is simple equal to the other proportion.

Thus, so far, we have two matrixes of the following form for each subject, i:

Trial t-1

		Word A	B	C	D	. . .	k . . .	KK	Sum
	A	1	1	0	1	. . .			3
	B	1	1	0	1	. . .			3
	C	0	0	1	0	. . .			1
W	D	1	1	0	1	. . .			3
o	.	.	.	.	.	. . .			.
r	.	.	.	.	.	. . .			.
d	.	.	.	.	.	. . .			.
	k								.
	.								.
	.								.
	.								.
	KK								.

Trial t

	Word A	B	C	D	. . .	k . . .	KK	Sum
A	1	1	0	0	. . .			2
B	1	1	0	0	. . .			2
C	0	0	1	0	. . .			1
D	0	0	0	1	. . .			1
.	.	.	.	.	. . .			.
.	.	.	.	.	. . .			.
.	.	.	.	.	. . .			.
k								.
.								.
.								.
.								.
KK								.

In the particular example depicted subject i formed a cluster of words A, B, and D on trial t-1, and for the portion of the table shown word C was not clustered with any other words,i.e., it was an isolate. In trial t the subject formed a cluster of words A and B, dropping word D from the cluster and making it an isolate. Word C remained an isolate. The index $R_{A(t-1)(t)}$ would then equal the square of (1·1 + 1·1 + 0·0 + 0·1 + ...) in the numerator; and the product of 3 times 2 in the denominator. The index of stability for word A in the sample would then be 4/(3·2) or 2/3. The index is the same for word B, equal to 1 for word C, and equal to 1/3 for word D.

II. Index of stability of organization for a cluster of words (cluster defined on trial t) from trial t-1 to trial t, for any given subject, i. The clusters on trial t are 1, 2, ...m, ... M. The index, $R_{m(t-1)(t)}$, is defined as:

(2) $R_{m(t-1)(t)} = \sum_{*} R_{k(t-1)(t)} / n_m$, where * means summation over all the words in the m^{th} cluster (defined on trial t), and n_m is simply the number of words in that cluster. Thus, $R_{m(t-1)(t)}$ is simply the arithmetic average of the $R_{k(t-1)(t)}$'s in the cluster.

III. Index of stability of organization from trial t-1 to trial t for any given subject, i. This is simply the arithmetic average of all KK of the $R_{k(t-1)(t)}$'s, i.e.,

(3) $R_{(t-1)(t)} = \sum_{k=A}^{KK} R_{k(t-1)(t)} / KK$

IV. Index of stability of organization from trial t-1 to trial t for any given set of subjects. This is simply the arithmetic average of the $R_{(t-1)(t)}$'s for all of the subjects in the set, and will be labelled $\overline{R}_{(t-1)(t)}$.

V. Index of stereotypy of clustering. This index reflects the degree to which a given word (e.g., word A) is clustered with the same words, across a set of Ss. Indexes I through IV are concerned with comparisons between trials t-1 and t, where the basic comparison is within a single S for a single word, and averages are then found across words, and then across Ss. The basic comparison for the index of stereotypy (Index V) is also for a single word, but it is not within a single S, and it involves only a single trial.

Thus, for a word (e.g., A) on a particular trial the index, r_A, is defined as follows:

Given the set of Ss (1, 2, 3, ..., i, ..., N), each S puts the word A into a cluster with from one (A is an isolated word) to n_{Ai} words (i.e., n_{Ai} - 1 other words). All possible pairs of Ss are considered, one pair at a time. For each pair of Ss (i and j) the index r_{Aij} is determined, where the index reflects the similarity or overlap of the words which each S associated with word A, i.e., put into a cluster with word A. The basic index is similar in form to the previous indexes in that the square of the count of the number of words both of a pair of Ss clustered with word A goes in the numerator; and the denominator is simply the number of words clustered with word A for one of the Ss (i), times the corresponding number for the other S (j) of the pair. The basic data may be entered into a matrix of the following form:

For Word A

		A	B	C	D	...	k	...	KK	n_{Ai}
S	1	1	1	0	1	. . .				3 = n_{A1}
u	2	1	1	1	1	. . .				4 = n_{A2}
b	3	1	0	1	0	. . .				2 = n_{A3}
j	4	1	0	0	0	. . .				1 = n_{A4}
e	.	.	.	.	.					.
c	.	.	.	.	.					.
t	.	.	.	.	.					.
s	i									n_{Ai}
	.									.
	.									.
	.									.
	j									n_{Aj}
	.									.
	.									.
	.									.
	N									n_{AN}
		—	—	—	—	— — —	—	— — —	—	—
N_{Ak}		N_{AA}	N_{AB}	N_{AC}	N_{AD}		N_{Ak}		N_{AKK}	$n_{A.}$
		4	2	2	2					

There would then be a similar matrix for each of the KK words.

For the Word A matrix: for each row (subject) there is a "1" in each column corresponding to words which that subject included in the same

category
subjective as word A; and a zero in the remaining cells of the row. In other words, let each cell entry, c_{Aik}, equal "1" if for subject i the kth word was included in the same subjective category with the word A. All other c's equal zero. The n_{Ai} are simply the sums of the entries in row i, and, for subject i, they are simply the number of words in the subjective category within which word A occurred. The N_{Ak} are the sums of the columns, and they are simply the number of subjects who included word k in the same subjective category as word A. The $n_{A\cdot}$ is simply the total of the ones in the entire table for word A, and it is the sum total of the number of words categorized with word A across the set of N subjects.

(Note: for a complete table for Word A, $N_{AA} = N$)

All possible pairs of subjects (rows) in the above Word A matrix are to be considered. There are, of course, (N/2)(N-1) different pairs. For each pair of subjects, i and j, define an index of common categorization as:

$$(4) \quad r_{Aij} = n_{Aij}^2 / (n_{Ai}) \cdot (n_{Aj}), \text{ where}$$

$$n_{Aij} = \sum_{k=1}^{KK} (c_{Aik}) \cdot (c_{Ajk})$$

n_{Ai} = Sum of row i, as defined in the Word A matrix above.

n_{Aj} = Sum of row j, as defined in the Word A matrix above.

The n_{Aij} are the counts of the words which both subjects i and j categorized with word A. Thus, the numerator for r_{A12} would be the square of the quantity $(1 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 + 1 \cdot 1)$, which equals 3^2. The denominator equals 3 times 4. And, $r_{A12} = 3/4$. All of the r_{Aij} for the example depicted in the Table for Word A are:

$$r_{A12} = 3^2/(3 \cdot 4) = 3/4$$

$$r_{A13} = 1^2/(3 \cdot 2) = 1/6$$

$$r_{A14} = 1^2/(3 \cdot 1) = 1/3$$

$$r_{A23} = 2^2/(4 \cdot 2) = 1/2$$

$$r_{A24} = 1^2/(4 \cdot 1) = 1/4$$

$$r_{A34} = 1^2/(2 \cdot 1) = 1/2$$

The index of stereotypy for Word A (for the set of subjects, N) is then defined as the arithmetic average of all of the r_{Aij}. For N subjects there are (N)(N-1)/2 different pairs of subjects, hence that many different r_{Aij}.

$$(5) \quad r_A = 2/(N)(N-1) \left[\sum_{\substack{i=1 \\ j>i}}^{N} \sum_{j=1}^{N} r_{Aij} \right].$$

For the above example $r_A = 2/(4 \cdot 3) \left[3/4 + 1/6 + 1/3 + 1/2 + 1/4 + 1/2\right] = (1/6) \cdot (30/12) = 0.4167$.

If only the first three of the subjects are considered the example would be:

$$r_{A12} = 3/4$$

$$r_{A13} = 1/6 \qquad r_A = 1/3 \left[(9 + 2 + 6)/12\right] = 0.472.$$

$$r_{A23} = 1/2$$

A computer program will be written to work with sets of 20 (or fewer) subjects. Two, or more, sets of subjects will provide reliability estimates for the index. The same thing that was done for Word A above is, of course, done for each of the KK Words, thus yielding an r_k for each of the KK words.

VI. Index of stereotypy of clustering for sets of words. This is simply the arithmetic average of the r_k for the words in the set. Sets may be defined by the experimenter (on apriori grounds, or on the basis of prior experimental evidence) to include any number of words from 2 through KK. The average across all KK words is, of course, the average stereotypy for the total set of stimulus words used.

VII. Index of common clustering for any specified pair of words. The intent is <u>not</u> to try to look at all possible pairs of words, but to use this index to examine only those particular pairs which are of special interest on apriori grounds or on the basis of prior experimental evidence. One obvious index which might be considered is simply the proportion of subjects, out of the set of N subjects, who include the particular pair of words in the same subjective category. For example consider the pair of words A and B. Utilizing the example depicted in the Table for Word A (the Table for Word B, depicted below, could also be used) this index would simply be N_{AB} / N_{AA} (which equals N_{BA} / N_{BB} from the Word B Table, below), or $2/4 = 1/2$. Thus, one half of the four subjects depicted in the example formed subjective categories such that words A and B were included within a single category. This simple index fails to take into account the sizes of the subjective categories in which words A and B are imbedded. If most subjects include these words in categories with many other words, i.e., large categories, then these two words would fall within the same category <u>by chance</u> more frequently than they would under the conditions in which the two words are usually imbedded in small categories. For this reason a somewhat more complex index will be used. It is similar in form to the previous indexes. It is defined as:

(6) $r_{AB} = (n_{AB})^2 / (n_{A\cdot})(n_{B\cdot})$, where

n_{AB} = the sum, across the set of N subjects, of the numbers of words in the categories within which the pair of words A and B may be found.

$n_{A\cdot}$ = the sum, across the set of N subjects, of the numbers of words in the categories within which the word A may be found

$n_{B\cdot}$ = the sum, across the set of N subjects, of the numbers of words in the categories within which the word B may be found.

In order to illustrate the computation of r_{AB} an illustrative table for Word B is also needed.

For Word B

		A	B	C	D	. . .	k	. . .	KK	n_{Bi}
s	1	1	1	0	1	. . .				3 = n_{B1}
u	2	1	1	1	1	. . .				4 = n_{B2}
b	3	0	1	0	0	. . .	1		1	3 = n_{B3}
j	4	0	1	0	0	. . .				1 = n_{B4}
e	.	.	.	.	.					.
c	.	.	.	.	.					.
t	.	.	.	.	.					.
s	i									n_{Bi}
	.									.
	.									.
	.									.
	j									n_{Bj}
	.									.
	.									.
	.									.
	N									n_{BN}
		—	—	—	—	———	—	———	—	———
N_{Bk}		N_{BA}	N_{BB}	N_{BC}	N_{BD}		N_{Bk}		N_{BKK}	$n_{B\cdot}$
		2	4	1	2		1		1	11

The quantity n_{AB} may now be obtained in any one of three ways. The first is from the Word A Table and is:

$$n_{AB} = \sum_{i=1}^{N} (c_{AiB}) \cdot (n_{Ai}).$$

The second is from the Word B Table and is:

$$n_{AB} = \sum_{i=1}^{N} (c_{BiA}) \cdot (n_{Bi}).$$

The third involves the Tables for both Word A and Word B, and is:

$$n_{AB} = \sum_{i=1}^{N} \sum_{k=A}^{KK} (c_{Aik}) \cdot (c_{Bik}).$$

All three yield identical answers. The quantities $n_{A\cdot}$ and $n_{B\cdot}$ are simply the sums of the ones in the Word A and Word B Tables, respectively, and are shown in the lower right hand corners of the Tables.

Using all four of the subjects depicted in the Tables:

$$r_{AB} = 7^2 \;/\; (10) \cdot (11) = 0.445.$$

Using only the first three subjects:

$$r_{AB} = 7^2 \;/\; (9) \cdot (10) = 0.545.$$

And using only the first two subjects:

$$r_{AB} = 7^2 \;/\; (7) \cdot (7) = 1.0.$$

VIII. Further computational illustrations. The computation of r_B, the index of stereotypy of clustering for Word B (Section V), is done from the example depicted in the Table for Word B as follows:

$$r_{B12} = 3^2 \;/\; (3 \cdot 4) = 3/4$$

$$r_{B13} = 1^2 \;/\; (3 \cdot 3) = 1/9$$

$$r_{B14} = 1^2 \;/\; (3 \cdot 1) = 1/3$$

$$r_{B23} = 1^2 \;/\; (4 \cdot 3) = 1/12$$

$$r_{B24} = 1^2 \;/\; (4 \cdot 1) = 1/4$$

$$r_{B34} = 1^2 \;/\; (3 \cdot 1) = 1/3$$

And $r_B = (2/12)\left[(27 + 4 + 12 + 3 + 9 + 12) \;/\; (36)\right] = 67/216 = 0.310.$

The computation of r_C, r_{AC}, and r_{BC} requires the use of the Table for Word C. An example Table is as follows:

For Word C

		A	B	C	D	. . .	k	. . .	KK	n_{Ci}
		0	0	1	0					1
s	1	1	1	1	1					4
u	2	1	0	1	0					2
b	3	0	0	1	0		1			2
j	4									
e	.									
c	.									
t	.									
s	i									
	.									
	.									
	.									
		2	1	4	1		1			9

$r_{C12} = 1^2 / (1 \cdot 4) = 1/4$

$r_{C13} = 1^2 / (1 \cdot 2) = 1/2$

$r_{C14} = 1^2 / (1 \cdot 2) = 1/2$

$r_{C23} = 2^2 / (4 \cdot 2) = 4/8 = 1/2$

$r_{C24} = 1^2 / (4 \cdot 2) = 1/8$

$r_{C34} = 1^2 / (2 \cdot 2) = 1/4$

And $r_C = (2/12)\ (2 + 4 + 4 + 4 + 1 + 2) / (8) = 17/48 = 0.354.$

Using all four of the subjects depicted in the Tables:

$r_{AC} = 6^2 / (9) \cdot (10) = 0.400.$

$r_{BC} = 4^2 / (9) \cdot (11) = 0.162.$

Using just the first three subjects depicted:

$r_{AC} = 6^2 / (7) \cdot (9) = 0.571.$

$r_{BC} = 4^2 / (7) \cdot (10) = 0.229.$

Using just the first two subjects depicted:

$r_{AC} = 4^2 / (5) \cdot (7) = 0.457.$

$r_{BC} = 4^2 / (5) \cdot (7) = 0.457.$

In summary of the samples depicted:

Subject	1	clustered	words	(ABD) (C)
"	2	"	"	(ABCD)
"	3	"	"	(AC) (BkK) (D?)
"	4	"	"	(A) (B) (Ck) (D?)

$r_A = 0.472.$

$r_B = 0.310.$

$r_C = 0.354.$

And for all four subjects:

$r_{AB} = 0.445.$

$r_{AC} = 0.400.$

$r_{BC} = 0.162.$

For just the first three subjects:

$r_{AB} = 0.545.$

$r_{AC} = 0.571.$

$r_{BC} = 0.229.$

For just the first two subjects:

$r_{AB} = 1.00.$

$r_{AC} = 0.457.$

$r_{BC} = 0.457.$

Derivation of the Bousfield formula as reported by Dallett (1964).

Repetitions expected by chance (RC) equals

$$RC = \frac{\sum_{i=1}^{k} (n_i)^2}{\sum_{i=1}^{k} n_i} - 1 = \frac{\sum_{i=1}^{k} (n_i)^2 - \sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} (n_i)(n_i - 1)}{\sum_{i=1}^{k} n_i}$$

where n_i is the number of words of the i^{th} category on a test, and k is the number of different categories on the test represented by at least one word.

Derivation:

1) For any n_i words the number of possible pairwise combinations (i.e., repetitions, or adjacencies) equals $(1/2)(n_i)(n_i - 1)$. It is simply the number of combinations of n_i things taken two at a time. For the i^{th} category it is the number of possible "successes" (i.e., adjacencies).

2) For k different categories the total number of possible "successes" is simply the sum of the number of successes for each of the categories, or

$$(1/2) \sum_{i=1}^{k} (n_i)(n_i - 1)$$

3) The total number of words on a test equals $\sum_{i=1}^{k} n_i = N$.

4) The total number of possible pairwise combinations of N words is $(1/2)(N)(N - 1)$.

5) Of the $(1/2)(N)(N - 1)$ total possible pairs there are

$(1/2) \sum_{i=1}^{k} (n_i)(n_i - 1)$ possible "successes". Dividing the number of possible successes by the total number of possible pairs yields the probability of any given pair, drawn at random, being a success.

Or,

$$\text{Prob. of any random pair being a "success"} = \frac{\sum_{i=1}^{k} (n_i)(n_i - 1)}{N\,(N - 1)}$$

6) In any ordered list of N words there are, in fact, N - 1 pairs (adjacencies).

7) Since any list of N words yields N - 1 pairs, and each pair has the same probability of being a "success," then the expected number of "successes" (i.e., chance adjacencies, or, RC) equals the number of pairs times the probability that a pair will be a "success". Thus,

$$RC = \frac{\sum_{i=1}^{k} (n_i)(n_i - 1)}{N\,(N - 1)} \times (N - 1) = \frac{\sum_{i=1}^{k} (n_i)(n_i - 1)}{N}$$

Since $N = \sum_{i=1}^{k} n_i$ (paragraph 3),

$$RC = \frac{\sum_{i=1}^{k} (n_i)(n_i - 1)}{\sum_{i=1}^{k} n_i}$$

Special case where all categories are, at most, two words big:

All tests have been scored for repetitions possible (RP), and for number correct (C). If all categories being examined for on the tests are, at most, two words big, then the RP and C measures for a given test may be used to calculate the RC measure for that test. Such is the case for the E-defined categories of Experiment Two. For every "repetition possible" on the test $n_i = 2$, or, $n_i = 2$ RP times. For every occurrence on the test of a single word of a pair, $n_i = 1$. For the present situation n_i can take only the values of one or two. When $n_i = 1$ it has no effect on the numerator of the formula for RC, which is $\Sigma(n_i)(n_i - 1)$, since $(n_i - 1) = 0$. Thus, the numerator is simply equal to $(2)(2 - 1)(RP)$. The denominator, Σn_i, is simply equal to the number of words correct, C. Hence,

$$RC = \frac{(2)(2-1)(RP)}{C} = \frac{2RP}{C}$$

This formula has some interesting, counter-intuitive properties. First, RC $\leq$ 1.0 regardless of the value of C, since 2RP $\leq$ C. In other words, no matter how many correct items there are on a test, the expected number of chance repetitions will be equal to, or less than, one. Second, if all words on the test are members of pairs of words, i.e., if 2RP = C, then RC = 1 regardless of the number of words on the test.

The quantity RO - RC for this special condition in which all E-defined categories are pairs is equal to

$$RO - RC = RO - 2(RP)/C$$

For the plots in Fig. 7 there are three E-defined pairs for each plotted point, hence RO for any one S can take the values 0, 1, 2, or 3. RC can vary between 0 and 1. Hence the quantity RO - RC must vary between 2 (for perfect clustering of all three pairs), through 0 (for the case where RO = RC, i.e., when both measures equal one or zero), to minus 1 (for the case where RO = 0 and all six words are present, but scattered throughout the test list, i.e., RP = 3, C = 6, and RC = 1).

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

(TOP)

ERIC REPORT RESUME

ERIC ACCESSION NO.	CLEARINGHOUSE ACCESSION NUMBER	RESUME DATE	P.A.	T.A.
		12 67		

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒

ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

001

100 TITLE
101 Final report: A Study of Methods of Organizing Learning.
102
103

200 PERSONAL AUTHOR(S):
Seibel, Robert

300 INSTITUTION (SOURCE):
College of Education
The Pennsylvania State University, University Park, Pa. 16802
SOURCE CODE

310 REPORT/SERIES NO.

320 OTHER SOURCE
SOURCE CODE

330 OTHER REPORT NO.

340 OTHER SOURCE
SOURCE CODE

350 OTHER REPORT NO.

400 PUB'L. DATE 12/67 — — CONTRACT/GRANT NUMBER OE-5-10-431

500 PAGINATION, ETC.
64p Project No. 5-8098-2-12-1
501 06/65 - 09/66

RETRIEVAL TERMS

600 learning
601 verbal
602 meaningful
603 subjective organization
604 strategy
605 cluster
606 free recall
experimental paradigm
study sheet
consistency of organization
stereotypy of organization
concept dominance
mutual relatedness
exhaustive categories

IDENTIFIERS

607

ABSTRACT

800 Given the task of learning a series of randomly ordered meaningful items, most
801 Ss impose some organization on the material during the process of learning. An
802 experimental paradigm is described which permits the objective scoring of each S's
803 subjective organization of the material on each learning trial. The central feature
804 of the paradigm is the utilization of an essentially unstructured study sheet which
805 each S prepares anew, for his own use, during each learning trial. New indices are
806 developed for measuring: a) the consistency of organization (independent of
807 sequential order) from one trial to the next; b) the stereotypy of organization
808 (i.e., degree of "sameness") across Ss on any one trial; and c) the extent to which
809 specific pairs of words are to be found in the same subjective categories for
810 different Ss. Three experiments are described. The extent to which Ss utilize E-
811 defined organization is evaluated. "Concept Dominance" is not persuasive, but high
812 degrees of "Mutual Relatedness" are. Much of the subjective organization behavior
813 depends on the total set of items involved. Overt organizing, and the utilization
814 of it, facilitates learning. There is no optimum size of a subjective cluster.
815 Consistency of word order from one test to the next does not increase with learning,
816 but consistency of organization does. The other two new indices reflect the Mutual
817 Relatedness variable. The third experiment indicates that in a mixed list of items,
818 exhaustive categories are learned faster than are exhaustive categories with one
819 item missing.
820
821
822